TEMPERATURE- AND PRESSURE-INDUCED PROTEIN DYNAMICS FROM MICROSECONDS TO MINUTES

BY

MAXIM B. PRIGOZHIN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemical Physics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Doctoral Committee:

Professor Martin Gruebele, Chair
Professor Zaida Luthey-Schulten
Professor Catherine J. Murphy
Professor Yann R. Chemla

ABSTRACT

For the past two decades, protein folding experiments have been speeding up from the second or millisecond time scale to the microsecond time scale, and full-atom simulations have been extended from the nanosecond to the microsecond and even millisecond time scale. Where the two meet, it is now possible to compare results directly, allowing force fields to be validated and refined, and allowing experimental data to be interpreted in atomistic detail.

This thesis describes recent experiments (and simulations) of fast protein folding ranging from microseconds to minutes using temperature and pressure as perturbation variables. Chapter 1 compares recent progress in the field of fast protein folding from experimental and computational sides. Chapters 2–5 are dedicated to unveiling the mechanism of folding of a model protein called λ-repressor fragment 6-85. Specifically, Chapter 2 describes an effort to identify the rate-limiting step of λ-repressor folding, Chapter 3 discusses potential origins of the slow (millisecond) phase in the folding of some λ-repressor mutants, Chapter 4 deals with λ-repressor refolding after a large ultrafast pressure jump, and Chapter 5 investigates the mechanism of λ-repressor folding monitored using multiple fluorescent probes to achieve better structural resolution of the folding process. Chapter 5 also compares λ-repressor folding triggered by temperature and pressure perturbations. Chapter 6 explores how the pressure-temperature phase-diagram of phosphoglycerate kinase is influenced by macromolecular crowding. In Chapter 7, an outreach project is described where simple mechanical and computer models of protein folding were used to teach students at the high school and undergraduate levels about scientific modeling and other basic concepts in physical chemistry and statistics.

The ultimate goal of these endeavors is to map out the energy landscapes of proteins and to generate "molecular movies", which reveal protein (mis)folding dynamics in atomistic detail. To this end, I have been striving

to refine experiments (and guide simulations) to provide better mechanistic detail and tackle the problems of multiple reaction coordinates, downhill folding, and complex underlying structure of unfolded or misfolded states.

To my parents

ACKNOWLEDGMENTS

"I realize how incredibly fortunate I have been [...] because so many people were prepared to put their faith in me [...] at a time when the evidence in favor of that proposition was nonexistent or perhaps even negative. I shall remain forever grateful."

Anthony J. Leggett

I would like to thank my Ph. D. adviser, Prof. Martin Gruebele, for allowing me to join his lab and for training me for the past four years. I learned *a lot*. Prof. Gruebele is a very eclectic scientist. Despite working primarily on protein folding, I learned quite a bit about quantum computation, nuclear fusion, ultrafast vibrational spectroscopy, glass dynamics probed by scanning tunneling microscopy, single-molecule absorption of carbon nanotubes, and even fish behavior (!) by simply attending group meetings and interacting with my labmates. Remarkably, Prof. Gruebele manages to explain all of these phenomena by invoking an appropriate energy landscape.

From a personality standpoint, it should suffice to say that Prof. Gruebele leads an incredible group of scientists who gravitate towards him for very good reasons. Group meetings have always been effervescent with questions, opinions and ideas thrown around uncontrollably – something I will definitely miss. It has been a pleasure and an honor to work with every single Gruebele group member who I have intersected with (listed in no particular order): Eduardo Berrios Rojas, Shu-Han Chao, Elaine Christman, Hannah Gelman, Kiran Girdhar, Duc Nguyen, Irisbel Guzman Sanchez, Minghao Guo, Lea Nienhaus, Anna Jean Wirth, Jonathan Tai, Jay Goodman, Max Platkov, Ionel Rata, Sharlene Denos, Praveen Chowdary, Krishnarjun Sarkar, Apratim Dhar, Sumit Ashtekar, Simon Ebbinghaus, Gregory Scott, Kapil Dave,

Drishti Guin, and Ruopei Feng. A few of them have now become my close friends.

I thank my committee members – Prof. Zaida Luthey-Schulten, Prof. Catherine Murphy, and Prof. Yann Chemla – for coaching me throughout my time in Illinois and putting up with my unsolicited progress updates. Other faculty members at Illinois and other universities also supported me at various stages of my graduate career (listed in no particular order): Prof. Nancy Makri (Illinois), Prof. Julio Fernandez (Columbia University), Prof. Vijay Pande (Stanford University), Prof. Charles Schroeder (Illinois) , Prof. Sua Myong (Illinois), Prof. Maria Spies (Illinois), Prof. Taekjip Ha (Illinois), Prof. Kenneth Suslick (Illinois), Prof. Steven Sligar (Illinois).

I am happy to say that today's protein folding science is a highly interdisciplinary effort with many papers contributed by international collaborative teams. Many of the projects described in this thesis would not have been possible without my collaborators in the US and Germany. I am grateful to Dr. Jed Pitera (IBM Almaden Research Center), Dr. William Swope (IBM Almaden Research Center), Prof. Vijay Pande (Stanford University), TJ Lane (Stanford University), Dr. Gregory Bowman (Stanford University), Dr. Shobhna Kapoor (TU Dortmund), Prof. Roland Winter (TU Dortmund), Prof. Klaus Schulten (Illinois), Dr. Yanxin Liu (Illinois), Dr. Taras Pogorelov (Illinois), Prof. Margaret Cheung (Houston University), Jianfa Chen (Houston University), Dr. Sharlene Denos (Illinois), and Prof. Gregory Scott (California Polytechnic State University).

Illinois has extremely supportive and dedicated staff in the School of Chemical Sciences. I would like to thank the members of the machine shop (Michael "Hodge" Harland, Brad Lutz, Rob Brown, Roger Smith, Dave Hire and Tom Wilson), members of the electronics and computer shop (Mike Hallock, John Rosheck and Ben Fisher) and the secretaries in the IMP Office (Beth Myler, Connie Knight, Theresa Struss and Karen Neumann) for providing outstanding service.

I would like to also thank several people who inspired me to do science in the first place. Going back to my high school years in Russia, I had a fantastic chemistry teacher, Svetlana Vladimirovna Gridneva. Something about the way Ms. Gridneva explained concepts resonated a lot with me. She was really able to breathe life into even the simplest unit-conversion problems and I am grateful to her for showing me that chemistry is much

more than balancing chemical equations. My undergraduate research adviser at the University of Toronto, Prof. Ulrich Krull, has been a huge inspiration. Prof. Krull's commendable breadth of expertise, extraordinary work ethic, and natural charisma defined my decision to pursue a doctoral degree. I am grateful to Prof. Krull for his support over the years. Taking a summer Research Opportunity Project course with Prof. Krull in 2006 (the summer after my freshman year) was easily the best decision I have ever made.

This thesis is dedicated to my parents to whom I am eternally grateful for going to the end of the earth to make sure that I get the most out of my life. Thanks to their perpetual efforts, I have always been able to do what I want and deliberately treat many things in life as peripheral to my education. I am also grateful to my brother, Eugene, for being there for me as long as I can remember: from helping me learn English at the age of 5 to regularly sending me funny images over the years that I have been abroad, which I only discover in the morning because of the time difference. Thanks must be extended to my girlfriend, Sarah, who is making me a better person every day by buffering fluctuations in my life, no matter how far from equilibrium, back to normal.

I am grateful to the Howard Hughes Medical Institute for awarding me the 2012 International Student Research Fellowship and to the National Science Foundation and the National Institutes of Health for funding my research projects.

TABLE OF CONTENTS

LIST OF TABLES xi

LIST OF FIGURES xiii

CHAPTER 1 MICROSECOND FOLDING EXPERIMENTS AND SIMULATIONS: A MATCH IS MADE 1
1.1 Introduction 1
1.2 Approaching the speed limit from above and below 6
1.3 Lowering the barrier to go downhill 14
1.4 How many reaction coordinates do we need? 22
1.5 Unfolded, trapped, misfolded 28
1.6 Challenges met and challenges to come 32

CHAPTER 2 REDUCING LAMBDA REPRESSOR TO THE CORE 35
2.1 Introduction 35
2.2 Experimental and Computational Methods 37
2.3 Results 40
2.4 Discussion and conclusion 49

CHAPTER 3 THE FAST AND THE SLOW: FOLDING AND TRAPPING OF λ_{6-85} 53
3.1 Introduction 53
3.2 Materials and methods 54
3.3 Results 57
3.4 Discussion and conclusion 63

CHAPTER 4 MISPLACED HELIX SLOWS DOWN ULTRAFAST PRESSURE-JUMP PROTEIN FOLDING 66
4.1 Introduction 66
4.2 Methods 68
4.3 Results 71
4.4 Discussion 81

CHAPTER 5 MULTIPROBE MAPPING REVEALS THE FAST FOLDING MECHANISM OF λ-REPRESSOR FRAGMENT 6–85 . 84
5.1 Introduction 84
5.2 Materials and methods 86
5.3 Results 89
5.4 Discussion and conclusion 96

CHAPTER 6 KINETIC AND THERMODYNAMICS EXPLORATION OF THE PRESSURE-TEMPERATURE PHASE DIAGRAM OF PHOSPHOGLYCERATE KINASE 97
6.1 Introduction 97
6.2 Experimental methods 98
6.3 Results 101
6.4 Discussion and conclusion 111

CHAPTER 7 MECHANICAL MODELING AND COMPUTER SIMULATION OF PROTEIN FOLDING 113
7.1 Introduction 113
7.2 The model 115
7.3 Mechanical prototype 117
7.4 Computer simulations 118
7.5 Summary 119

APPENDIX A SUPPLEMENTARY MATERIAL FOR "REDUCING LAMBDA REPRESSOR TO THE CORE" 122
A.1 Explicit solvent replica-exchange molecular dynamics of λ_{6-85} 122

APPENDIX B SUPPLEMENTARY MATERIAL FOR "THE FAST AND THE SLOW: FOLDING AND TRAPPING OF λ_{6-85}" 127
B.1 Thermal titrations of λ_{6-85} mutants monitored by fluorescence and CD 127
B.2 Temperature jump kinetics of λ_{6-85} mutants spanning 5 ms . . 135
B.3 Global model fit 141

APPENDIX C SUPPLEMENTARY MATERIAL FOR "MISPLACED HELIX SLOWS DOWN ULTRAFAST PRESSURE-JUMP PROTEIN FOLDING" 143
C.1 Kinetic data 143
C.2 Guanidine titrations 143
C.3 Temperature titrations 143
C.4 Pressure titrations 147
C.5 Molecular dynamics simulations 151

APPENDIX D SUPPLEMENTARY MATERIAL FOR "MULTIPROBE MAPPING REVEALS THE FAST FOLDING MECHANISM OF λ-REPRESSOR FRAGMENT 6–85" 157

APPENDIX E SUPPLEMENTARY MATERIAL FOR "KINETIC AND THERMODYNAMICS EXPLORATION OF THE PRESSURE-TEMPERATURE PHASE DIAGRAM OF PHOSPHOGLYCERATE KINASE" 167

APPENDIX F SUPPLEMENTARY MATERIAL FOR "MECHANICAL MODELING AND COMPUTER SIMULATION OF PROTEIN FOLDING" 171
F.1 Materials 171
F.2 Mechanical model 171
F.3 Computer simulation 180
F.4 Folding thermodynamics and kinetics 186
F.5 Protein folding in the cell 197

REFERENCES 204

LIST OF TABLES

2.1 Residue ranges and sequences of lambda repressor helix variants 41

2.2 Twenty nine truncated variants, including two wild type linkers for reference. Average folded and unfolded energies, energy difference per atom, and average C_α-RMSD for helices 1 and 4 at 300 K are shown. Reported uncertainty estimates are one standard deviation. Italic entries were expressed and studied experimentally. 43

2.3 Fitting results for the thermodynamic titration experiments ($\pm$ errors are two standard deviations of the mean) 46

3.1 λ-repressor nomenclature, mutations from wild type, and melting temperatures (rounded to the nearest °C) as determined by fluorescence wavelength shifts 58

3.2 Ratios of amplitudes of slow- and fast-folding phases in λ-repressor mutants having different stabilities. 61

A.1 Average radii of gyration and uncertainty (in Å) over the last 1.25 ns of each folded simulation 124

A.2 Average radii of gyration and uncertainty (in Å) over the last 1.25 ns of each unfolded simulation 125

A.3 Sequences for each lambda repressor variant 126

B.1 Thermodynamic midpoints for unfolding of λ-repressor mutants as measured by tryptophan fluorescence and CD spectroscopy. Here SVD 1 represents the thermodynamic midpoints determined by the decrease in the intensity and SVD 2 represents the thermodynamic midpoints determined by the wavelenght shift of the W22 fluorescence spectrum or CD spectrum of λ-repressor mutants. SVD 2 could not be extracted from λD14A and λnQ33Y data because the shift of CD spectrum was not large for these mutants. The reported values are based on changes in the intensity of the CD spectrum at 222 nm. 133

B.2 Thermodynamic parameters fitted or held fixed for Scheme S1(A) fitted to the λQ33Y data in Figure 3.3A (blue curves, main paper). 142

D.1 Fitting parameters for the equilibrium temperature melt probed by fluorescence. 157
D.2 Fitting parameters for the equilibrium temperature melt probed by circular dichroism 162
D.3 Fitting parameters for the temperature jumps probed by tryptophan fluorescence lifetime. T^* is the temperature to which the temperature jump was done. 165
D.4 Fitting parameters for the temperature jumps probed by tryptophan fluorescence lifetime. T^* is the temperature to which the temperature jump was done, which in each case is several degrees below T_m for each protein. 165
D.5 Fitting parameters for the temperature jumps probed by tryptophan fluorescence lifetime for λ_{13} at 54 °C. 165
D.6 Fitting parameters for the equilibrium temperature melt probed by circular dichroism 166
D.7 Fitting parameters for the equilibrium temperature melt probed by fluorescence. Concentration of GuHCl is 1.2 M. . . 166
D.8 Fitting parameters for temperature and pressure jumps to the same final condition for λ_{13} and λ_{32}: [GuHCl] = 1.2 M, T = 23 °C. For λ_{12}, [GuHCl] = 2.4 M. 166

E.1 Fitting parameters for the equilibrium pressure melts probed by fluorescence and analyzed by mean wavelength at various temperatures. 167
E.2 Fitting parameters for the equilibrium temperature melts probed by fluorescence and analyzed by mean wavelength at 1 and 500 bar. 167
E.3 Fitting parameters for the equilibrium pressure melts at 100 mg/mL ficoll probed by fluorescence and analyzed by mean wavelength at various temperatures. 168
E.4 Fitting parameters for the equilibrium temperature melts at 100 mg/mL ficoll probed by fluorescence and analyzed by mean wavelength at 1 and 500 bar. 169
E.5 Fitting parameters for the equilibrium pressure melts at 100 mg/mL ficoll probed by fluorescence and analyzed by mean wavelength at various temperatures. 170

F.1 End-to-end distances for the unfolded protein. 175
F.2 End-to-end distances for the folded protein. 175
F.3 End-to-end distances for the unfolded protein. 178
F.4 End-to-end distances for the unfolded protein. 179

LIST OF FIGURES

1.1 Relation between populations and free energies of states. (A) The example free energy landscape has three wells. N is the compact helical native state, U is the expanded unfolded state containing residual helix, and B is the compact misfolded state rich in β-sheet. The sequence of states N–B–U or B–U–N depends on the choice of reaction coordinate. If the N–B barrier is high, there is a preferred path B–U–N. Therefore if only one reaction coordinate is chosen to describe the system, helix content would be better than radius of gyration. (B) The population in U is very small compared to B or N because U lies several RT above B and N. Clearly, this does not mean that U is not involved in the interconversion from B to N. By tuning the solvent condition or mutating the protein sequence in (C), it is possible to bring U to lower free energy so its population will be larger than B.[1] The nature of the populated non-native states preceding folding, and the actual paths taken, are sensitive to initial conditions.[2] A dominant pathway is observed because population is exponentially sensitive to small changes of the free energy (Boltzmann factor). Two pathways are rarely going to lie at exactly the same free energy, although it has been observed.[3, 4] 5

1.2 Folding from activated to downhill, using WW domain experiments and simulations as an example. (A and C) The bottom shows schematic free energy profiles[5, 6] along the C_α-RMSD reaction coordinate at temperatures below and above the melting temperature T_m. Protein populations in folded/unfolded minima (orange) and near the barrier (green) are shown as dots. The top (A) shows two probe signal profiles. They increase monotonically with reaction coordinate, and can therefore serve as experimental reaction coordinates. Different probes progress differently as the protein folds,[6] yielding probe-dependent kinetics particularly when the barrier is low and the protein population samples the transition region (vertical dashed line). (B) On the molecular time scale τ_m (vertical dashed line) the rate coefficient is not a 'rate constant', but depends on time[7]: pre-activated population (green in C) reacts promptly, much faster than population that needs to be activated (orange in C). Only later does the rate coefficient 'settle down'. 9

1.2 (Previous page.) (D) In the example, at $T > T_m$ the pre-activated population is negligible, and exponential-decay kinetics with a 'slow' time constant ?a is observed. At $T < T_m$, the pre-activated population is large, and a prompt phase precedes the 'slow' exponential-decay kinetics. The insets in (D) show corresponding single molecule traces: for a high barrier ($T > T_m$ in the example), the activated protein (green), is sampled rarely. For a low barrier ($T < T_m$), the activated protein is sampled frequently. In essence, there are always pre-activated proteins that fold promptly downhill at the 'speed limit.' If the barrier is large, this population is unobservably small due to the Boltzmann factor. If the barrier is small, the population becomes easy to observe.[8] The terms 'molecular time', 'speed limit', 'transition state transit time', 'downhill folding time' refer to the same time scale, but are not identical. Also, the decay in (D) at $t < \tau_m$ is not necessarily an exponential with time constant τ_m,[9] although frequently fitted as such.[6, 8, 10] 10

1.3 Experimental signatures of downhill folding upon protein stabilization: only the fastest-folding, most stable mutants of lambda repressor fragment have a significant population undergoing prompt reaction (the molecular phase shown in Figure 1.2D). (A) The measured molecular phase amplitude increases smoothly when the activated rate k_a increases towards the molecular rate $k_m \approx 1\mu s^{-1}$ (Yang & Gruebele[11]; Liu & Gruebele (1) and (2)[11, 12]; Ma & Gruebele[9]), as predicted when the free energy barrier approaches RT (downhill folding).[13] 15

1.3 (Previous page.) (B) The kinetics of mutants that are relatively unstable can be fitted by slow single-exponential kinetics upon temperature jumps (orange area); their activation barrier is too high to carry a measurable pre-activated population. Mutants that have $T_m > 60$ °C show an additional fast molecular phase (triangles) because their barrier is low enough so there is a promptly reacting (downhill folding) protein population (C) On the left: the normalized enthalpy of the polypeptide chain generally decreases when the configurational entropy s_c decreases: as favorable contacts are made, the polypeptide chain moves less freely. Folding is 'downhill' in enthalpy (folding is an exothermic reaction), resulting in an 'enthalpy funnel', but this is not what is meant by downhill folding. On the right: the free energy G can be computed from the enthalpy and entropy as a function of an arbitrarily chosen reaction coordinate x by evaluating H and S at x and averaging over all other orthogonal coordinates. 'x' could be the radius of gyration, distance between two FRET labels, *etc.*, and is normalized from -1 (unfolded) to 1 (native) here. Of course, a carefully chosen set of coordinates $x, y, ...$ provides a more complete description of a reaction as complicated as folding than just a single coordinate x. The free energy has a barrier (orange) when the enthalpy does not funnel the protein towards the native state efficiently enough to offset the decreasing entropy (orange funnel on the left). The free energy is downhill (green) when the exothermicity of the reaction is sufficient to offset the loss of entropy everywhere along the reaction coordinate (green funnel on the left). The protein then folds with the molecular rate k_m instead of the slower rate k_a (black circle in (B)). In intermediate cases both rates can be measured simultaneously (triangles + circles in (B) connected by a dot, or $T < T_m$ trace in Figure 1.2D), allowing an absolute determination of the free energy barrier height[6, 8–12]. 16

1.4 Systematic benchmark studies of empirical force fields to improve their performance in folding simulations are now more frequently appearing in the literature[14–20]. (A) Various force fields were tested by Lindorff-Larsen *et al.*[21] and a score based on the performance of the force field against the chosen model systems was devised such that the lower score indicates better agreement with experimental data. The plot shows the improvement of force fields over time. (B) Piana *et al.*[22] used four different force fields to fold villin headpiece. Although all simulations arrived at the correct native state, the folding mechanism depended on the force field used. The panel shows that the flux through different reaction pathways (123 *etc.* is the order in which the three helices of villin headpiece form) is a function of the force field that was used. 21

1.5 α_3D is a designed protein[23] for which a one-dimensional reaction coordinate cannot explain the probe-dependent kinetics using a reasonable diffusion coefficient[24]. (A) Structure of α_3D from PDB (2A3D). (B) A contour plot of the optimal two-dimensional free energy surface for α_3D calculated using Langevin dynamics as described by Scott *et al.*[25] (C) Kinetic rates measured with infrared absorption spectroscopy are shown as black dots, fluorescence spectroscopy as red dots. The dashed (one-dimensional model) and solid (two-dimensional model) lines represent the fits of the data using Langevin dynamics, but the one-dimensional fit requires an unrealistic diffusion coefficient; diffusion coefficients of incorrect magnitude or with unusual coordinate dependence are a warning sign that the model underestimates the dimensionality of the dynamics. (D) Time traces of Q and C_α-RMSD for α_3D from Lindorff-Larsen *et al.* show strong correlation, but are not equivalent[5].Q is the fraction of long-range native contacts. For the quantitative definition of Q see page 3 of Supplemental Materials in ref. [5]. 27

1.6 Various conformations of villin headpiece drawn from a simulation by Beauchamp *et al.*[26] The left-most conformation is the native state. Three structures in the middle broadly represent the unfolded state ranging from the partially disordered conformations that still resemble the native state to the significantly extended conformations with low residual secondary content. The right-most structure is a misfolded trap. Such traps lie off the predominant folding pathway.[27] 30

1.7 At the top: contact map of protein L.[28] The blue squares indicate residues that are in contact in the native state. The structure of the native state (PDB: 2PTL) is shown at the bottom-left of the figure. Data for the contact map were generated using the CMA server at http://ligin.weizmann.ac.il/cma/). The grey scale contour map shows the average distances (0–2 nm) between all pairs of residues at 300 K as simulated by Voelz *et al.*[28] The red dots indicate the contact map for one of the conformations of the unfolded state from the simulation. The structure of this state is shown at the bottom-right of the figure. Clearly this unfolded state is not a random coil but a partially structured non-native conformation. 31

2.1 Model structure of λ^*_{6-85} based on the X-ray crystal structure of a closely related lambda repressor fragment PDB 3KZ3,(5, 20) and molecular dynamics models of the structures for $\lambda_{blue1-3}$ computed in the present work. A van der Waals surface is also shown for each model, along with their sequences. 44

2.2 Circular dichroism spectra (A) and thermal denaturations for $\lambda_{blue1-3}$. Thermodynamic fits are shown as black lines in the bottom plot, with values for the fitting parameters in Table 2.3. 45

2.3 Fluorescence intensity-detected thermal denaturation traces for $\lambda_{blue1-3}$. Thermodynamic fits are shown as black lines in the bottom plot, with values for the fitting parameters in Table 2.3. The inset shows the λ_{blue1} transition with the average overall baseline subtracted to reveal the transition more clearly. 47

2.4 CD spectra and thermal melts of N terminal and C terminal fragments of λ_{blue1}. The mixture of both peptides in A simply produces an average of the two individual fragment melting curves. The spectra in B show much more random coil content (negative CD below 210 nm) than full λ_{blue1} in Figure 2.2. 48

2.5 Temperature jump experiment: 0.5 ms of data containing over 30,000 tryptophan fluorescence decays were collected (the blue band shows fluorescence intensity), with a detail of the raw data near the T-jump shown below. Raman scatter from the infrared pump beam obscures a few fluorescence decays, setting the dead time of ≈50 ns. The top graph shows the fluorescence lifetime analysis $\chi(t)$ for λ_{blue1} (15 μs fitted response) and for tryptophan (instantaneous response). The 15 μs relaxation phase is also visible in the intensity data near $t = 0$. 50

3.1 Optical components of the nanosecond laser temperature jump instrument. 56
3.2 Thermal denaturation of the λ-repressor mutants in Table 3.1. The singular value decomposition (SVD) shown here represents the wavelength shift of the W22 fluorescence spectrum (raw data are given in the Appendix). The inset shows W22 in the crystal structure of λQ33Y.[29] 58
3.3 T-jump kinetics of λ-repressor mutants detected by tryptophan fluorescence decay, with double-exponential fits in black. The global model fits are shown as blue curves in (A). Traces were normalized from 1 (fast fluorescence decay) to 0 (slow decay). The fast phase (gray areas) was investigated previously,[30–32] so the ms scale is emphasized here. 60
3.4 Dependence of the A_{slow}/A_{fast} ratio on temperature with and without GuHCl. (open square: fixed point.) 62
3.5 Folding of λ_{6-85} mutants along reaction coordinates related to compactness (R_g) and non-native secondary structure content. (A) Free energy landscape. The folding process per se from an extensively unfolded state E (or U) to N is fast. E can also convert slowly to compact non-native states T and T′. Suggested structures are from refs [33] and [29] (N). (B) Kinetic scheme. T and T′ interconvert indirectly through E, which is in rapid equilibrium with N. . . 64

4.1 Pressure denaturation of λ^*YA probed by fluorescence spectroscopy of a 200-μM sample (A–C) and fluorescence lifetime analysis of a 300-μM sample (D). (A) Fluorescence spectra of λ^*YA in guanidine (pH 7) at 100-bar intervals from 1 to 2,500 bar (rainbow gradient). The basis spectrum of the folded state is shown in black, and that of the fully unfolded state is shown in purple (Appendix C). a.u., arbitrary units. (B) Fluorescence peak shift (centroid) as a function of pressure. λ^*YA in 2.4 M GuHCl shows a much larger 0.0032-nm/bar shift than λ^*YA in buffer (native state model) or NATA (denatured state model). (C) Fraction folded was calculated by fitting the spectra in A to a linear combination of the folded and unfolded basis spectrum (two-state model; Appendix C); at 1,200 bar (initial condition for P-jumps), ≈40% of the protein is unfolded. The crystallographic structure of λ^*YA obtained from the PDB (ID code 3KZ3) is shown. (D) Scaled fluorescence lifetime change relative to NATA (1 at 1 bar, 0 at 1,200 bar). NATA and λ^*YA in 0 M GuHCl lifetimes decrease linearly with pressure, whereas λ^*YA in 2.4 M GuHCl shows the onset of pressure denaturation (χ for proteins was shifted up by +3 because NATA has a much longer lifetime; Appendix C, Figure C.4). 72

4.2 P-jump instrument. The sample is pipetted into a dimple in a sapphire cube. The dimple is covered with mylar-coated aluminum foil and pressurized by pumping ethanol into a pressure fitting. A current burst into a copper electrode bursts the upper steel membrane and releases the pressure. Sample fluorescence is excited by a 280-nm pulsed laser every 12.5 ns and is collimated by a UV light guide onto a photomultiplier. The digitized raw data consist of a train of fluorescence decays, whose lifetime and intensity monitor the refolding of the sample after the sudden P-drop at $t = 0$. 73

4.3 P-jumps (300-μM sample) and T-jumps (200-μM sample) of λ^*YA and NATA, probed by tryptophan fluorescence decays. Tryptophan lifetime change was normalized for NATA so that $\chi = 0$ corresponds to the decay lifetime before the jump (1,200 bar) and $\chi(t) = 1$ corresponds to the decay lifetime 5 ms after the jump (1 bar). The rest of the jumps were analyzed using the lifetime decays from the P-jump of NATA for direct comparison. Solid black curves are the double-exponential fits of the data with relaxation times $\tau_f = 3.8 \pm 0.4$ μs and $\tau_s = 1.4 \pm 0.2$ ms for the P-jump and $\tau_f = 63 \pm 2$ μs and $\tau_s = 2.17 \pm 0.02$ ms for the T-jump. 75

4.4 Equilibrium denaturation of λ^*YA by pressure and temperature, probed by IR spectroscopy (1.7-mM sample). (A) IR absorbance spectra of λ^*YA in the amide I$'$ region measured at 295 K. Triangles indicate 1 bar, and circles indicate 13.9 kbar. These spectra were used as basis functions for the analysis of the entire pressure denaturation curve (Methods and Appendix C). The IR absorbance spectrum of λ^*YA in the amide I$'$ region measured at 1 bar and 368 K is shown as a gray dashed line. (B) Denaturation of λ^*YA as a function of pressure [$\chi(P) = 1$ means the 1-bar basis function contributes 100% of the signal, $\chi(P) = 0$ means the 13.9-kbar basis function contributes all the signal]. A thermodynamic two-state fit of the data is shown as a solid black curve, and the error bars are the residuals (Appendix C, Figure C.5). The midpoint of pressure denaturation, P_m, is equal to 6.0 $\pm$ 0.2 kbar in the absence of denaturant. 77

4.5 CHARMM27 simulation of λ^*YA during P-drop. (Upper) Structures from the two trajectories. The high-pressure simulations start with 1 μs at 325 K and 5 kbar (blue zone), followed by a 0.15-μs P-drop to 1 bar (white zone). (Lower) Refolding (8.85 μs) was simulated at 1 bar and 325 K. Central carbon atom root mean square displacement values were calculated relative to the crystal structure (PDB ID code 3KZ3) [34]. The fraction of residues in α (gray) and β (red) conformations is shown. R_{gyr} is the unsolvated radius of gyration. The native mean values (green solid lines, except red for β-fraction) are from a 150-ns equilibrium simulation of the native structure at T = 325 K and P = 1 bar. 80

4.6 Residue-specific α-helical propensity of the simulations in Figure 4.5 (black, first simulation; green, second simulation). The helical percentage was defined as the time percentage each residue spent in α-helical conformation during the last 8 μs of refolding simulation. The secondary structure of the crystal structure is shown as a color-coded background, and the sequence at the top, together with the red arrows, highlights turn/coil residues with >75% helix content in both simulations. 81

5.1 New lambda repressor mutants. (a) Crystal structure of lambda repressor mutant Q33Y, Y22W, G46A, G48A (PDB ID: 3KZ3). Helices 1–5 of the bundle are highlighted in red, yellow, green, blue and purple, respectively. W22 and Y33 are shown as stick models in red and yellow, respectively. The other mutations (F51, A62, and L69) are shown in gray. The Van der Waals surface of the protein is shown in transparent gray. (b) The sequences of the designed mutants. The mutants were named as λ_{WY}, where W is the helix in which tryptophan is located and Y is the helix in which tyrosine is located. λ_{42} is the L69W mutant, λ_{42b} is the A62W mutant. Alpha-helical regions are shown at the bottom. (c) Distances between W and Y for each mutant obtained from 100 ns molecular dynamics relaxations. Error bars represent standard deviations. 90

5.2 Stability of the nee lambda mutants. (a) Spectral mean of fluorescence spectra of the new lambda mutants as a function of temperature. λ_{12} is shown in gray for comparison. Solid lines are two-state thermodynamic fits. (b) Mean residue ellipticity at 222 nm as a function of temperature for the brew mutants. λ_{12} is shown in gray for comparison. Solid lines are two-state thermodynamic fits. λ_{42} did not show a cooperative transition and was not fitted. (c) Integrated fluorescence intensity normalized to begin at 1 and end at 0 for the new lambda repressor mutants. λ_{12} is shown in gray for comparison. Solid lines are two-state thermodynamic fits. Note that an increase in fluorescence intensity upon denaturation is observed for all mutants except λ_{42}, indicated a lack of W–Y interaction in the folded state. (d) Experssion efficiency and melting temperatures of the new lambda mutants. Acrylamide gels indicate good expression of λ_{13} and λ_{32}, satisfactory expression of λ_{42} and poor expression of λ_{42b} and λ_{41}. 91

5.3 Kinetics of lambda repressor. Fluorescence lifetime normalized from 1 to 0 in arbitrary units and plotted against the logarithmic time axis. $\lambda_{blue1}(14)$ is from Prigozhin et al. (2011). λ_{12} is from Prigozhin et al. (2011). The solid black lines are single-exponential fits of the data: $I(t) = e^{-t/\tau}$. Gaussian profiles at the bottom correspond to the four decays above. The profiles are centered at the fitted τ_{obs} and have the width equal to twice the standard deviation. 92
5.4 Stability of the new lambda repressor mutants as a function of guanidine concentration. (a) Spectral mean of fluorescence spectra of the new lambda mutants as a function of guanidine concentration. λ_{12} is shown in gray for comparison. Solid lines are two-state thermodynamic fits. (b) Integrated fluorescence intensity normalized to begin at 0 and end at 1 for the new lambda repressor mutants. λ_{12} is shown in gray for comparison. Solid lines are two-state thermodynamic fits. (c) Mean residue ellipticity at 222 nm as a function of guanidine concentration for the brew mutants. λ_{12} is shown in gray for comparison. Solid lines are two-state thermodynamic fits. 94
5.5 Robustness of the folding mechanism of λ_{6-85} with respect to the perturbation variable. (a) Spectral mean for fluorescence spectra of lambda measured in 1.2 M guanidine at 1 bar as a function of temperature (circles). Solid lines are two-state thermodynamics fits. Dashed lines are the melts at 0 M guanidine for each mutant shown for reference. (b) T-jump kinetics of the two mutants from 14 °C to 23 °C at 1.2 M guanidine at 1 bar. (c) Spectral mean for fluorescence spectra of lambda measured in 1.2 M guanidine at 23 °C as a function of pressure (circles). Dashed lines are the melts at 0 M guanidine for each mutant shown for reference. (d) P-jump kinetics of the two mutants from 1600 bar to 1 bar in 1.2 M guanidine at 23 °C. 95

6.1 Behavior of phosphoglycerate kinase in the pressure-temperature plane. (a) Crystal structure of wild type phosphoglycerate kinase (PDB ID: 1QPG). The backbone of the protein is shown in blue, the Van der Waals surface is shown in transparent gray and Y122 (W122 in the destabilized variant) of domain 1 is shown in yellow. (b) Fluorescence spectra of PGK taken in increments of 100 bar at 23 °C from 1 bar to 2500 bar. (c) Spectral mean of the spectra (similar to the ones plotted in (b) at various temperatures. The solid lines represent two- or three-state thermodynamic fits. Note that three thermodynamic states are observed at 9, 15, and 23 °C. Profiles at 30 and 36 °C show two-state behavior. The protein is denatured at 44 °C. (d) Spectral mean of temperature melts of PGK at 1 and 500 bar. The solid lines represent two-state thermodynamic fits. (e) Temperature dependence of the negative change in volume between the native state and the intermediate state (circles, transition 1), and between the intermediate state and the denatured state (triangles, transition 2). (f) Pressure temperature-phase diagram of PGK. Circles represent midpoint pressures (at 9, 15, 23, 30, 36 °C) or temperatures (at 1 and 500 bar) for transition 1. Triangles represent midpoint pressures for transition 2. Solid elliptical curve going through the circles is a fit representing the $\Delta G = 0$ curve. A curve going through the triangles is a guide for the eye. (g) Free energy change profiles for transition 1 (circles) and transition 2 (triangles) as a function of temperature at various pressures. Solid lines going through the circles are the van't Hoff equation fits. Dashed lines are the extrapolated van't Hoff fits with a constant allowing for the negative apex of the parabola. 102

6.2 Pressure-induced denaturation of phosphoglycerate kinase at various ficoll 70 concentrations. (a) Spectral mean of the spectra (similar to the ones plotted in 6.1b and 6.3b) at various ficoll concentrations. T = 23 °C. The solid lines represent two- or three-state thermodynamic fits. Note that three thermodynamic states are observed only when no ficoll is present. Profiles at 50, 100, 150, and 200 mg/mL all show two-state behavior. (b) Midpoint pressures obtained from the fits to data shown in panel (a) plotted as a function of ficoll concentration. Only midpoint pressures of transition 1 are plotted. Transition 2 only appears when no ficoll is present. The solid line is a linear fit of the data. (c) Negative change in volume between the native state and the intermediate state (transition 1) as a function of ficoll concentration. (d) Free energy change profiles for transition 1 as a function of ficoll concentration at various pressures. 106

6.3 Behavior of phosphoglycerate kinase in the pressure-temperature plane in the presence of 100 mg/mL ficoll 70. (a) Ficoll 70 is a polymer with molecular weight ≈ 70 kDa. (b) Fluorescence spectra of PGK crowded by ficoll taken in increments of 100 bar at 23 °C from 1 bar to 2500 bar. (c) Spectral mean of the spectra (similar to the ones plotted in (b) at various temperatures ([ficoll] = 100 mg/mL). The solid lines represent two- or three-state thermodynamic fits. Note that three thermodynamic states are observed at 9, 15 °C, not at 23 °C as was the case without ficoll. Profiles at 23, 30 and 36 °C show two-state behavior. The protein is denatured at 44 °C. (d) Spectral mean of temperature melts of PGK at 1 and 500 bar ([ficoll] = 100 mg/mL). The solid lines represent two-state thermodynamic fits. (e) Temperature dependence of the negative change in volume between the native state and the intermediate state (circles, transition 1), and between the intermediate state and the denatured state (triangles, transition 2). Black – no ficoll, blue – 100 mg/mL ficoll. (f) Pressure temperature-phase diagram of PGK. Black – no ficoll, blue – 100 mg/mL ficoll. Circles represent midpoint pressures (at 9, 15, 23, 30, 36 °C) or temperatures (at 1 and 500 bar) for transition 1. Triangles represent midpoint pressures for transition 2. Note that transition 2 is not observed in the presence of ficoll at 23 °C. Solid elliptical curves going through the circles are the fits representing the $\Delta G = 0$ curves. Curves going through the triangles are guides for the eye. (g) Free energy change profiles for transition 1 as a function of temperature at various pressures with (blue) and without (black) ficoll. 107

6.4 Pressure melts of PGK with and without ficoll probed by infrared spectroscopy in the amide I′ region. (a) Infrared spectra of PGK crowded by ficoll (100 mg/mL) at 23 °C from 1 bar to 9000 bar. (b) Infrared spectra of PGK under pressure without ficoll (black) were linearly normalized to go from 1 to 0 as described in the Data Analysis section. Spectra of PGK crowded by 100 mg/mL ficoll were then analyzed using the basis functions of spectra without ficoll to obtain data points in blue. The solid line going through the blue circles is a three-state thermodynamic fit. 109

6.5 Pressure-jump kinetics of phosphoglycerate kinase with and without ficoll 70. (a) Pressure-jumps of PGK from 1600 bar to 1 bar. T = 23 °C. The panel on the left shows a microsecond-to-millisecond time scale experiment, 1 bin = 1.25 μs. The panel on the right shows a millisecond-to-second time scale experiment, 1 bin = 10 ms. The data are normalized from 1 to 0 based on the recorded fluorescence lifetime (see Data Analysis). (b) Pressure-jumps of PGK from 900 to 1000 bar at various ficoll 70 concentrations. The data points were generated using the spectral mean analysis and normalized from 1 to 0. T = 23 °C. Solid lines are single-exponential (for [ficoll] = 0 and 50 mg/mL) or double-exponential (for [ficoll] = 100, 150 and 200 mg/mL) fits. (c) Observed time constants, τ_{obs}, from the fits in (b). Dashed lines are guides for the eye. 110

7.1 Equations of state for gases as a function of simplicity and accuracy of each model. The simplest model is the ideal gas law (on the left), while one of the more accurate ones is the Benedict-Webb-Rubin-Starling (BWRS) equation of state (on the right). The two extremes are balanced by introducing two (instead of twelve in the BWRS equation!) extra parameters into the ideal gas law, which results in a formula that we know as a Van der Waals model (in the middle). 114

7.2 Models of proteins as a function of simplicity and accuracy of each model. (A) A rigid crystallographic state, which is commonly used to visualize a protein. If this state was the only one that actually existed in reality, the free energy surface for the protein would look like a narrow well with the minimum in free energy corresponding to the crystallographic structure. (B) A protein under solution conditions that strongly favor the folded state would largely only populate the native conformation, but thermal fluctuations would cause it to "breathe" deviating slightly from the crystallographic state. This behavior can be conceptualized as a broadening of the free energy well in (A). (C) A protein under solvent conditions where the folded state is preferred but the conformational space can be explored beyond the folded well to populate unfolded structures. An extra well is added on the free energy surface for the unfolded state with a barrier separating the two states. (D) The same protein can exhibit a multitude of stabilities and folding mechanisms in a heterogeneous environment of the living cell. Such locally specified parameters as viscosity, pH, ionic strength and crowding can influence the exact shape of the free energy landscape of the protein. 116

7.3 A two-dimensional mechanical prototype of a 10 amino-acid protein made with paper clips. The folded protein contains a rigid zigzag created by soldering 4 joints on one side of the chain. (B) A Styrofoam box is used to confine the protein prototype to a rectangular area to approximate cellular crowding. 118

7.4 User interface of the computer program used to model the behavior of the prototypical protein. As inputs the program requires the number of samplings, the probability to switch to the unfolded state if the current state is folded and the conjugate probability going in the other direction. The outputs include a series of folded and/or unfolded protein conformations, a histogram of their end-to-end distances, the center of mass of the histogram, a plot of dwell times in each state as well as the equilibrium constant and the folding rate. 120

B.1 Thermal titrations of λQ33Y monitored by fluorescence (A–C) and CD (D–F). (A) Spectra collected at ∼3 °C increments from 20 °C (black curve) to 94 °C (blue curve). Integrated intensity decreases with increasing temperature. (B) SVD basis vectors. (C) SVD trends with corresponding two-state thermodynamic fits. (D) Spectra collected at 3 °C increments from 20 °C (black curve) to 98 °C (blue curve). After the melt was done, the temperature was again decreased to 20 °C to show reversibility (green curve). (E) SVD basis vectors. (F) SVD trends with corresponding two-state thermodynamic fits. 128
B.2 Thermal titrations of λsQ33Y monitored by fluorescence (A–C) and CD (D–F). (A) Spectra collected at ∼3 °C increments from 20 °C (black curve) to 94 °C (green curve). Integrated intensity decreases with increasing temperature. (B) SVD basis vectors. (C) SVD trends with corresponding two-state thermodynamic fits. (D) Spectra collected at 3 °C increments from 20 °C (black curve) to 98 °C (green curve). After the melt was done, the temperature was again decreased to 20 °C to show reversibility (red curve). (E) SVD basis vectors. (F) SVD trends with corresponding two-state thermodynamic fits. 129
B.3 Thermal titrations of λD14A monitored by fluorescence (A–C) and CD (D–F). (A) Spectra collected at ∼3 °C increments from 20 °C (black curve) to 94 °C (blue curve). Integrated intensity decreases with increasing temperature. (B) SVD basis vectors. (C) SVD trends with corresponding two-state thermodynamic fits. (D) Spectra collected at 3 °C increments from 20 °C (black curve) to 98 °C (blue curve). After the melt was done, the temperature was again decreased to 20 °C to show reversibility (red curve). (E) SVD basis vectors. (F) SVD trends with corresponding two-state thermodynamic fits. 130
B.4 Thermal titrations of λnQ33Y monitored by fluorescence (A–C) and CD (D–F). (A) Spectra collected at ∼3 °C increments from 20 °C (red curve) to 94 °C (blue curve). Integrated intensity decreases with increasing temperature. (B) SVD basis vectors. (C) SVD trends with corresponding two-state thermodynamic fits. (D) Spectra collected at 3 °C increments from 20 °C (red curve) to 98 °C (blue curve). After the melt was done, the temperature was again decreased to 20 °C to show reversibility (green curve). (E) SVD basis vectors. (F) SVD trends with corresponding two-state thermodynamic fits. 131

B.5 Thermal titrations of λsA49G monitored by fluorescence (A–C) and CD (D–F). (A) Spectra collected at ∼3 °C increments from 20 °C (black curve) to 94 °C (blue curve). Integrated intensity decreases with increasing temperature. (B) SVD basis vectors. (C) SVD trends with corresponding two-state thermodynamic fits. (D) Spectra collected at 3 °C increments from 20 °C (black curve) to 98 °C (blue curve). After the melt was done, the temperature was again decreased to 20 °C to show reversibility (red curve). (E) SVD basis vectors. (F) SVD trends with corresponding two-state thermodynamic fits. 132
B.6 Thermal denaturation of λ-repressor mutants. SVD shown here represents the decrease in the intensity of the W22 fluorescence spectrum. The inset shows the crystal structure of λQ33Y. Note the qualitatively different behavior of the λs49G mutant (yellow), which lacks the Q33Y mutation. All other mutants show a relative increase of fluorescence upon denaturation, superimposed on an overall decrease of tryptophan fluorescence intensity at higher temperature, as observed also for tryptophan itself. 134
B.7 Optical components of the nanosecond laser temperature jump instrument. 135
B.8 5 °C temperature jumps of 300 μM *L*-tryptophan solution to the final temperature of 45 °C at incident power of the UV beam from Ti:Sapphire laser equal to 8 mW and 4 mW. The traces are normalized such that the average of the first and last 100 data points of $\chi(t)$ are set to 1 and 0, respectively. At 4 mW excitation power, fluorescence lifetime of tryptophan increases slightly after the jump due to slow recooling of the sample to the equilibrium temperature. At 8 mW, the cooling trend is overpowered by the opposing effect of irreversible photobleaching, which reduces the fluorescence lifetime of tryptophan over time. 136

B.9 Correction of kinetics data for the effects of cooling and bleaching. The overall trends observed remained the same even without this correction. Top left: Kinetic trace of λD14A to the final temperature of 60 °C is shown in red. A similar set of data is shown for λD14A at 60 °C but without a jump (hence no folding phase) in blue. Kinetic trace in red contains the effects of cooling and bleaching along with the folding signature of λD14A but the trace in blue only contains the effect of λD14A bleaching at 60 °C, which is subtracted out from the red trace to yield the green time trajectory. The latter then contains the effects of cooling and folding of λD14A at 60 °C. Bottom left: The same procedure was followed to account for the bleaching effect in *L*-tryptophan at 60 °C. Center right: Assuming cooling of λD14A and Trp at 60 °C after a T-jump are similar we subtract the Trp time trace that contains only the cooling effect from that of λD14A that has both the cooling and the folding signature. The resulting time trace (black) can then be fitted to a double exponential (orange curve). 137
B.10 Temperature jumps of the more stable mutants of λ-repressor: λQ33Y (A and B) and λsQ33Y (C and D) at 0 M (A and C) and 0.5 M (B and D) GuHCl. Black curves represent the bi-exponential fits of the data. The temperature jumps occurred at $t = 0$. Temperatures after the jump are indicated above the corresponding time traces. 138
B.11 Temperature jumps of *L*-tryptophan (A–C) and λD14A (D–F), a mutant of intermediate thermal stability, at 0 M (A and D), 0.5 M (B and E), and 1 M (C and F) GuHCl. Black curves represent the bi-exponential fits of the data. The temperature jumps occurred at $t = 0$. Temperatures after the jump are indicated above the corresponding time traces. 139
B.12 Temperature jumps of the less stable mutants of λ-repressor: λnQ33Y (A and B) and λsA49G (C and D) at 0 M (A and C) and 0.5 M (B and D) GuHCl. Orange (λnQ33Y) and black (λsA49G) curves represent the bi-exponential fits of the data. The temperature jumps occurred at $t = 0$. Temperatures after the jump are indicated above the corresponding time traces. 140

B.13 Three-state kinetic schemes that were compatible with the experimental data regardless of realistic tryptophan lifetimes of the states N, E or T, or whether they are consistent with the computational results from Bowman *et al*. Of the models shown, only (A) and (B) produced reasonable relative tryptophan lifetimes (smaller = more quenched for N than for extensively unfolded state, T either quenched or similar to E). Model (A) is also consistent with the simulation results and with the notion that the native state can not directly convert to a compact beta sheet state without first going through an extensively unfolded state U⇔E. 141

C.1 The refolding kinetic data in Figure 4.3 of the main text is an average of 3 pressure drops of λ*YA in 2.4 M GuHCl. Here the data are analyzed with respect to the Trp fluorescence lifteime decay of the protein ($\chi = 0$ means the data is perfectly fitted by the decay f_1 before the jump, $\chi = 1$ means the data is perfectly fitted by the decay f_2 at 5 ms, other values of χ mean the data is fitted by a linear combination $(1 - \chi)f_1 + \chi f_2$. On the right the NATA P-jump (also analyzed with its own f_1 and f_2 functions) is highlighted near $t = 0$, showing the time resolution obtained for 1200 bar P-jumps, which is somewhat lower than for the 2400 bar P-jumps we reported previously (Dumont *et al*. reference in main text). 144

C.2 GuHCl titration of λ*YA showing 2.4 M onset of denaturation. The titration was monitored by fluorescence (A–C, 5 μM sample) and CD (D–F, 25 μM sample). (A, D) Spectra collected at 0.2 M increments of GuHCl from 0 M (green curve) to 5 M (magenta curve). (B, E) SVD basis vectors. (C, F) SVD trends with corresponding two-state thermodynamic fits. Vertical dashed lines highlight the 2.4 M GuHCl point. 145

C.3 Reversibility of λ*YA unfolding after a temperature denaturation at different protein concentrations as monitored by CD spectroscopy. (A–C) 2.5 μM, (D–F) 50 μM, (G–I) 100 μM. (A, D, G) CD spectra of λ*YA. Spectra in red follow the signal as the temperature increases, grey spectra represent the signal at the highest temperature, and spectra in blue show the signal as the temperature is decreasing back down. (B, E, H) The first SVD basis vectors for the corresponding data sets in A, D and G, respectively. (C, F, I) The first SVD basis vector trends color coded the same way as in panels A, D and G for the corresponding data sets in A, D and G, respectively. 146
C.4 Fluorescence decays of λ*YA (left) and NATA (right) before the pressure-jump (function f_1, P = 1.2 kbar, blue curve), right after the pressure-jump (P = 1 bar, black curve), and at the end of the time trace (function f_2, P = 1 bar, red curve). The decays shown here correspond to an average of 100 decays. Note that NATA has a significantly longer fluorescence lifetime than Trp in the protein, for this reason χ was shifted up by 3 units in for the protein lifetime data in Figure 4.1D. 147
C.5 Residual of fitting the IR-detected pressure denaturation to the basis functions in Figure 4.4A to obtain the denaturation curve in Figure 4.4B. (A) Residuals were obtained by subtracting the fitted spectra from the data, integrating over the wavelengths, and dividing the resulting value by the integral of the data spectrum. (B) The raw data as well as the fit for the largest residual value (solid black circle in (A)). 148
C.6 Reversibility of NATA and λ*YA with and without GuHCl unfolding after a pressure denaturation as monitored by fluorescence spectroscopy. (A–C) NATA, (D–F) λ*YAwithout GuHCl, (G–I) λ*YA with GuHCl. (A, D, G) Fluorescence spectra of NATA, λ*YA without GuHCl, and λ*YA with GuHCl, respectively. Spectra in red follow the signal as the pressure increases, and spectra in blue show the signal as the pressure is decreasing back down. (B, E, H) The first (black) and the second (green) SVD basis vectors for the corresponding data sets in A, D and G, respectively. (C, F, I) The first (black) and the second (green) SVD basis vector trends for the corresponding data sets in A, D and G, respectively. The pressure increase is shown as a solid line and the pressure decrease is shown as a dashed line. . . . 149

C.7 FTIR spectroscopy of λ*YA with 0.5 M GuDCl. (A) In red: FTIR spectra of λ*YA shifting towards lower energy from 1 bar to 13.9 kbar without GuDCl. Basis vectors that were used for the folded (black) and unfolded (yellow) protein are also shown. In blue: FTIR spectra of 0.5 M GuDCl from shifting towards higher energy from 1 bar to 14.2 kbar. In green: FTIR spectra of λ*YA with 0.5 M GuDCl. (B) Basis spectra used for χ-analysis of the green spectra in panel A. Black and yellow spectra correspond to λ*YA at 1 bar and 13.9 kbar, respectively, as shown in panel A. The blue spectra were obtained by doing χ-analysis on the GuDCl spectra (blue spectra in panel A) separately, fitting the resulting $\chi(P)$ to a straight line and using that linear fit to arrive at GuDCl spectra at the pressures at which λ*YA with 0.5 M GuDCl (green spectra in panel A) was measured. (C) χ-coefficients for the GuDCl basis spectrum (blue), basis spectrum for the folded λ*YA (black), and the basis spectrum for the unfolded λ*YA (yellow). As expected, χ-coefficient for GuDCl stays flat, the folded population decreases, while the unfoled population increases as the pressure increases. 150

C.8 Generation of λ-repressor P-T unfolded initial states using three different force fields: CHARMM27[35], CHARMM36[36], and AMBER299SB[37]. C_α-RMSD values have been calculated relative to the crystal structure 3KZ3[38]. α/β content is the fraction of residues that are in the α-helical conformation (black dot) and β-sheet conformation (red dot). R_{gyr} refers to the radius of gyration. The average native values, calculated from a 150-ns equilibrium simulation of the native structure at T = 325 K and P = 1 bar, are shown as green solid lines. The unfolding simulation started from the crystal structure λ-repressor fragment at T = 325 K and the pressure gradually increased in 0.15 μs from 1 bar to 5 kbar as indicated by the colored background; then the temperature was increased to 525 K and maintained for another 0.15 μs to completely unfold the protein while keeping the pressure at 5 kbar. The resulting λ-repressor conformation at the end of the pressurization and heating steps are shown at the top. Charmm27 has residual helical structure, Charmm36 and Amber299SB do not. 152

C.9 Secondary structure throughout the two refolding trajectories using CHARMM27 force eld[35]. The secondary structure of the crystal structure is shown at the beginning as a reference. Color coding for secondary structure is shown at the top. 153
C.10 Like Figure 4.5, simulation of λ*YA during pressure drop using CHARMM36[36] and AMBER299SB[37] force fields. The initial states for these simulations are the corresponding final states form Figure C.8. Top: structures from the two trajectories. The high pressure simulations start with 1 μs at 325 K and 5 kbar (blue zone), followed by 0.15 μs pressure drop to 1 bar (white zone). 8.85 μs of refolding were simulated at 1 bar and 325 K. C_α-root mean square deviations were calculated relative to the crystal structure 3KZ3[38]. α/β-content is the fraction of residues that are in the α-helical and β-sheet conformation, respectively. R_{gyr} is the unsolvated radius of gyration. The native values are dened by the mean values (green solid line) of the 150-ns equilibrium simulation of the native structure at T = 325 K and P = 1 bar. Note that the CHARMM36 structure (R_{gyr}) fluctuates more than any other we simulated. 154
C.11 Simulation of λ*YA refolding using CHARMM27[35] and CHARMM22*[39] force elds. Both simulations start from the AMBER299SB force eld in Figure C.3. C_α-root mean square deviations were calculated relative to the crystal structure 3KZ3[38]. α/β-content is the fraction of residues that are in the α-helical and β-sheet conformation, respectively. R_{gyr} is the radius of gyration. The native values are dened by the mean values (green solid line) of the 150-ns equilibrium simulation of the native structure at T = 325K and P = 1 bar. 155
C.12 Residue-specic α-helical propensity of the refolding simulation using CHARMM22*[39] in Figure C.11. The helical percentage was dened as the time percentage each residue spent in a α-helical conformation during the last 8 μs of refolding simulation. The secondary structure of the crystal structure is shown as a color-coded background, and the sequence at the top highlights turn/coil residues with >75% helix content in the simulations using CHARMM27 force eld (see Figure 4.6). Two of the loop regions identified in Charmm27 simulations (in red) also have helix propensity in the Charmm22* simulations. 156

D.1 Temperature melts of the new lambda repressor mutants probed by fluorescence spectroscopy. (a–c) Raw data for λ_{13}, λ_{32}, and λ_{42}, respectively. The spectra were taken from 23 °C to 95 °C in 3 °C increments. (d) Singular value decomposition basis functions 1. (e) Singular value decomposition basis function 1 trends. Note that both λ_{13} and λ_{32} show an increase in fluorescence intensity upon unfolding, while λ_{42} does not. (d) Singular value decomposition basis functions 2. (e) Singular value decomposition basis function 2 trends. 158

D.2 Temperature melts of the new lambda repressor mutants probed by circular dichroism spectroscopy. (a–c) Raw data for λ_{13}, λ_{32}, and λ_{42}, respectively. The spectra were taken from 5 °C to 95 °C in 3 °C increments. (d) Singular value decomposition basis functions 1. (e) Singular value decomposition basis function 1 trends. Note that both λ_{13} and λ_{32} show a cooperative transition upon unfolding, while λ_{42} does not. (d) Singular value decomposition basis functions 2. (e) Singular value decomposition basis function 2 trends. . . . 159

D.3 Kinetic traces of lambda repressor mutants observed after a 9 °C temperature jump. (a) Kinetic traces of $lambda_{13}$ at various temperatures. (b) Same data as in (a) plotted against logarithmic time axis. (c) Kinetic traces of $lambda_{32}$ at various temperatures. (d) Same data as in (c) plotted against logarithmic time axis. (e) Same data as in the kinetics figure in Chapter 4 but plotted against the linear timescale. (f) Concentration dependence of $lambda_{13}$ kinetics. A slight slowdown is observed when the concentration is doubled from 25 μM to 50 μM. 160

D.4 Guanidine hydrochloride titrations of λ_{13} and λ_{32} probed by fluorescence spectroscopy. (a, b) Raw data for λ_{13} and λ_{32}, respectively. The spectra were taken from 0 M GuHCl to 5 M GuHCl in 0.2 M GuHCl increments. (c) Singular value decomposition basis functions 1 for both data sets. (d) Singular value decomposition basis function 1 trends. (e) Singular value decomposition basis functions 2 for both data sets. (f) Singular value decomposition basis function 2 trends. 161

D.5 Guanidine hydrochloride titrations of λ_{13} and λ_{32} probed by circular dichroism spectroscopy. (a, b) Raw data for λ_{13} and λ_{32}, respectively. The spectra were taken from 0 M GuHCl to 5 M GuHCl in 0.2 M GuHCl increments. (c) Singular value decomposition basis functions 1 for both data sets. (d) Singular value decomposition basis function 1 trends. 162
D.6 Pressure melts of λ_{13} and λ_{32} with and without GuHCl probed by fluorescence spectroscopy. (a–d) Raw data for λ_{13} (a, b) and λ_{32} (c, d) with (b,d) and without (a, c) GuHCl. The spectra were taken from 1 bar to 2500 bar in 100 bar increments. (e) Singular value decomposition basis functions 2 for all four data sets. (f) Singular value decomposition basis function 2 trends. 163
D.7 Temperature melts of λ_{13} and λ_{32} in 1.2 m GuHCl probed by fluorescence spectroscopy. (a, b) Raw data for λ_{13} and λ_{32}, respectively. The spectra were taken in 3 °C increments. (c) Singular value decomposition basis functions 1. (d) Singular value decomposition basis function 1 trends. (e) Singular value decomposition basis functions 2. (f) Singular value decomposition basis function 2 trends. 164

E.1 Temperature- and pressure-induced thermodynamics of the N-terminal domain of PGK. (a) N-terminal domain of PGK. (b) Circular dichroism spectrum of the N-terminal domain of PGK showing substantial secondary structure content. (c, g) Fluorescence spectra of the N-terminal domain of PGK under temperature and pressure, respectively. (d, h) Spectral mean of spectra in (c) and (g), respectively. 168
E.2 Hysteresis in the pressure thermodynamics of PGK. (a) Change in fluorescence spectra on the way up and on the way down at 1000 bar without ficoll. (b) Change in fluorescence spectra on the way up and on the way down at 1000 bar in 100 mg/mL ficoll. (c) Hysteresis in the pressure thermodynamics of PGK with (blue) and without (black) ficoll on the way up (circles) and on the way down (triangles). (d) Kinetics at 1000 bar on the way up (circles) and on the way down (triangles) with (blue) and without (black) ficoll. 169

E.3 Concentration dependence of pressure-induced kinetics of PGK. (a, b) Downward pressure jumps at 33 μM and 66 μM, respectively. Spectra in blue were taken at 1 bar before the pressure was increased. Spectra in green are the spectra taken right before the pressure was dropped. Spectra going from black to yellow represent a pressure-jump from 1600 bar to 1 bar. (c, d) Upward pressure jumps at 33 μM and 66 μM, respectively. Spectra in blue were taken at 1 bar before the pressure was increased. Spectra in green are the spectra taken right before the pressure was increased from 900 bar to 1000 bar. Spectra going from black to yellow represent a pressure-jump from 900 bar to 1000 bar. (e) Spectral mean representation of data in (a–d) showing all the data points when the pressure was increased before the pressure-jumps. Pressure-jumps occurred at $t = 0$. Solid lines are double-exponential fits. 170

F.1 A cartoon of a 10 amino acid model protein. 172
F.2 Histogram 1. Bin = 1 cm. 176
F.3 Histogram 2. Bin = 5 cm. 177
F.4 Typical student data for Histogram 1. 177
F.5 Typical student data for Histogram 2. 178
F.6 Graphical user interface of the simulation program 180
F.7 Protein folding reaction of interest 187
F.8 Free energy landscape of the reaction of interest 187
F.9 The free energy landscape shown in Figure 8 is in gray. The free energy landscape where $K_{eq} = 1$ and $\Delta G = 0$ is in black. 194
F.10 The free energy landscape shown in Figure 8 is in gray. The desired free energy landscape is in black. 195

CHAPTER 1

MICROSECOND FOLDING EXPERIMENTS AND SIMULATIONS: A MATCH IS MADE

1.1 Introduction*

Folding science is a vast field, drawing upon its tools from chemistry, physics, computational science, molecular biology, bioengineering and many other areas. It covers proteins, but also other biomolecules with varied structures such as nucleic acids. This Chapter specifically takes a look at how microsecond protein folding experiments and microsecond simulations (either single- or multi-trajectory) have come together to sharpen our understanding of how small globular proteins fold. The era of folding science was ushered in with the structures and thermodynamic principles that began to be revealed by experiments in the 1960s.[40, 41], Computational protein folding began to develop soon thereafter,[42, 43] progressing from model systems[44, 45] to explicit solvent simulations of multiple complete folding/unfolding events.[5] The era of direct comparison between protein folding experiment and simulation began in the late 1990s when both converged on the microsecond time scale.[46–54] This was made possible by advances in computing power and force fields to push simulation towards microseconds from the bottom up[55–58] and by the development of ever faster folders and new initiation/detection techniques to push experiments towards microseconds from the top down.[59–61]

1.1.1 Experiments

Fast protein folding is often studied by subjecting the sample to a rapid perturbation of a state variable (e.g. temperature, pressure, pH, concentration

* This chapter is partially reproduced from Maxim B. Prigozhin and Martin Gruebele, Physical Chemistry Chemical Physics, 15, 3372–3388, 2013, DOI: 10.1039/C3CP43992E

of denaturant) and monitoring the evolution of the system to a new thermodynamically stable state. The perturbation must be much faster than the rate at which the system is able to respond. A common example is the temperature jump (T-jump). In this method the temperature of protein solution is suddenly (ns time scale) increased by ≈5–15 K and the subsequent unfolding of the proteins is probed by fluorescence of the amino acid tryptophan. The resulting signal is representative of the behavior of an ensemble of molecules.[61]

In the simplest "two state" case, often assumed as a default for small, fast-folding proteins, the detected response would be a single exponential decay, due to the interconversion of protein populations between the initial and final thermodynamic state. The observed time constant ?a is the inverse of the activated rate coefficient k_a, which is the sum of the folding and unfolding rates because both reactions are monitored simultaneously at the ensemble level.[62]

A similar result can be obtained at the single-molecule level.[60] If the system is ergodic, monitoring a single protein molecule jumping from one state to the other will yield histograms of dwell times in each of the two states. For two states connected by a large activation barrier, the histograms will follow exponential distributions, yielding time constants for folding and unfolding. In relation to fast protein folding, single-molecule experiments retain their advantages in terms of the ability to differentiate between multiple states and their connectivities more easily. Ensemble experiments have the advantage that better (<1 μs) time resolution already can be achieved, which is naturally important for fast protein folding studies.

Both types of experiments can detect more complex mechanisms, revealing additional states occupied during the folding process. When the barriers connecting these states are $\gg$RT, we label them as "intermediates". When the barriers connecting the states approach RT, we label them as landscape roughness and speak of a protein folding downhill. This is not an either-or distinction, but a gray scale of more-or-less separable time scales (larger barriers = longer time scales of interconversion). Very recently, fast folding experiments and simulations are making forays into this area,[63, 64] which was already studied extensively by observing slowly interconverting "intermediates" as long ago as the 70s.[65]

1.1.2 Simulations

Protein folding can be simulated in full detail using powerful computers.[66] In these simulations each atom of a protein is subjected to a potential, *i.e.* force field, which is determined by the bonds that the atom participates in, the angles that these bonds make with the neighboring atoms and the nonbonding pair interactions with every other atom in the simulation.[56] Three major approaches have emerged to extract long timescale information: long single trajectories with recurring folding/unfolding events observed for a single protein,[5] multiple shorter trajectories that can be stitched together into an overall picture of folding,[67, 68] and replica or sequential sampling methods that provide thermodynamic information from multiple 'copies' of a protein.[69, 70] The simulation methods thus span a similar range of philosophies as the experiments.

In the near future, we will undoubtedly see the success of integrating experiment, simulation and theory spread from the current microsecond folders to millisecond and slower folders, and to other complex molecular systems that are currently too expensive to model, such as membrane proteins. These advances will be driven by further improvement of force fields and computational speed, and by new experimental techniques that provide time-resolved yet structurally rich information about folding, both *in vitro* and in more realistic solvation environments such as live cells. Structurally detailed comparison of computation and experiment in different folding environments brings about its own challenges. First, tying together multiple reaction coordinates from experiment and modeling into a coherent whole at an unprecedented level of structural detail is the next-level unresolved problem *in vitro* and in silico. Second, comparing results *in vitro* and *in vivo* will reveal how cells and organisms modulate the energy landscape to control the action of biomolecules and utilize both frequent (high population) and rare (high energy) dynamic phenomena optimally.

1.1.3 Energy landscapes and pathways

Simulation and experiment are connected by statistical mechanical models, often referred to as the 'energy landscape picture'.[71–76] This energy landscape involves reducing most of the solvent degrees of freedom and most of

the protein degrees of freedom (e.g. dihedral angles of the amino acids) to a small number of coordinates or 'order parameters'. There has been a lot of debate in the literature whether proteins fold down 'one path' *vs.* 'on the landscape'. Fast folding experiments and simulations show that both pictures are useful, depending on whether population or energy is considered to be the most important variable.

From a biological perspective, population is the most important variable. Proteins are classified into coarse-grained populations, such as 'folded' or 'unfolded'. Systems biology deals with such populations at the level of "A interacts with B". This view is indispensable when the complexity at the molecular level is too great to treat in full detail. For example, there is much experimental and computational evidence that fast-folding globular proteins tend to populate a predominant path en route to the folded state.[77, 78]

From a physics perspective, energy is the most important variable. Protein folding is therefore characterized by an energy landscape. The folded ensemble lies low on this landscape, the unfolded ensemble is high-lying in energy. There is much experimental and computational evidence from fast-folding globular proteins that many distinct low-energy states and paths for folding exist on such landscapes, and that different paths can be selected by protein engineering or by choice of solvent conditions.[2, 24, 79] Some paths temporarily 'park' proteins in traps,[3] while others make the folding process more robust by providing alternate routes.[4]

Both views have their place in folding science. They are mutually consistent because population P is very sensitive to energy E. The two are related by the exponential Boltzmann factor, $P \propto e^{-E/RT}$ (Figure 1.1). For example, an alternative path just 3 RT above the minimum energy path contributes less than 5% to the observed population. That is not to say the 3 RT path is unimportant. First, in another mutant of the protein, it actually can be the lowest-lying path taken to the native state.[2] Second, evolution is partly based on phenotypic selection of mutants, and low-lying paths are thus input for evolutionary variation of proteins.[80] Third, higher energy conformations visited by the protein during folding or unfolding may be important for signaling, catalysis or other functions by facilitating conformational rearrangement. The protein may access such conformations only rarely, but they are key to biological function.[81, 82]

Fast folding proteins make it easier to see non-native conformations ex-

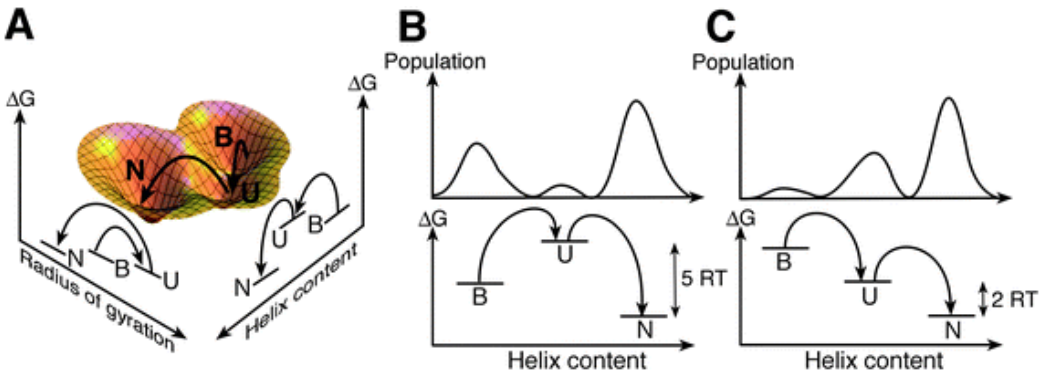


Figure 1.1: Relation between populations and free energies of states. (A) The example free energy landscape has three wells. N is the compact helical native state, U is the expanded unfolded state containing residual helix, and B is the compact misfolded state rich in β-sheet. The sequence of states N–B–U or B–U–N depends on the choice of reaction coordinate. If the N–B barrier is high, there is a preferred path B–U–N. Therefore if only one reaction coordinate is chosen to describe the system, helix content would be better than radius of gyration. (B) The population in U is very small compared to B or N because U lies several RT above B and N. Clearly, this does not mean that U is not involved in the interconversion from B to N. By tuning the solvent condition or mutating the protein sequence in (C), it is possible to bring U to lower free energy so its population will be larger than B.[1] The nature of the populated non-native states preceding folding, and the actual paths taken, are sensitive to initial conditions.[2] A dominant pathway is observed because population is exponentially sensitive to small changes of the free energy (Boltzmann factor). Two pathways are rarely going to lie at exactly the same free energy, although it has been observed.[3, 4]

perimentally because speed implies smaller energy barriers and gaps, and therefore larger Boltzmann factors. Low-lying paths also make it easier for simulations to sample rare but important folding events (*e.g.* transition state passage[6]) for direct comparison with experiments. Hence there has been a strong interest in studying small, fast-folding proteins *in vitro* and in silico. Experiment and simulation have now converged on the microsecond time scale, making fully quantitative comparisons possible for the first time. This quantitative convergence on the microsecond folding time scale is the focus of this Chapter. The following sections are organized by a 'Key concept', followed by conceptual 'Elaboration', and finally 'Support' from the fast folding experimental and computational literature.

1.2 Approaching the speed limit from above and below

1.2.1 Key concept

The 'speed limit' of protein folding is the fastest time it would take a protein of certain size to fold on an energy landscape with the lowest possible activation barrier(s). The speed limit of folding for single-domain proteins of ≈1 μs was established by T-jump and single-molecule fluorescence experiments as well as molecular dynamics (MD) simulations.

1.2.2 Elaboration

The chemical reactions of small organic molecules occur in picoseconds or faster, about the time required to make or break a chemical bond.[83] Although the solvent is intimately involved in these reactions through polarization, viscosity and other properties, we can think of small molecule reactions by using the gas phase formalism of transition state theory because the solvent is not altered in a fundamental way during a reaction. Although the crossing of the free energy barrier by an individual molecule takes only picoseconds, the activated rate coefficient $k_a = \tau_a^{-1}$ can be very small at room temperature: barriers often exceed $E_a = 100$ kJ mol^{-1}, so the probability of reaching the activation energy, proportional to the Boltzmann factor $e^{-E_a/RT}$,

is very small. Usually the fast barrier crossing dynamics and the slow activated reaction kinetics occur on well-separated time scales, the molecular time scale τ_m and the activated time scale τ_a:

$$\tau_a = \tau_m e^{+E_a/RT} \gg \tau_m \tag{1.1}$$

Not so for protein folding. Multiple weak interactions such as hydrogen bond formation, hydrophobic exclusion of water molecules from the protein core, or salt bridge formation occur during folding. The polypeptide chain moves through a solvent that fully participates in these interactions and contributes a large fraction of the folding free energy through its own reorganization. Thus it is almost surprising that folding can be treated as a series of ordinary chemical reaction steps, using a one-dimensional picture analogous to equation 1.1, such as Kramers' theory, simply by substituting "G_a" instead of "E_a" in equation 1.1 .

The separation of time scales between molecular time and activated time is certainly not as large for folding as it is for most small molecule reactions, nor is coarse-graining to one reaction coordinate as accurate. Nonetheless this picture is useful as a starting point. When $G_a \to 0$ and the two time scales meet at the 'speed limit',[59] the Kramers analog of equation 1.1 no longer provides a satisfactory description of the folding process. There are many reasons why τ_m in equation 1.1 should be picked much slower for proteins than for small molecules. For example, folding requires large-amplitude polypeptide chain motions through a viscous solvent. Also, polypeptides have many coordinates, so numerous unproductive motions orthogonal to the chosen reaction coordinate decrease the apparent diffusion coefficient of the protein along that reaction coordinate.

τ_m is the minimum time that must elapse before an activated reaction can be described by equation 1.1 with a constant rate coefficient $k_a = \tau_a^{-1}$. To visualize this statement, consider Figure 1.2C. The barrier of the upper free energy profile is large, so a negligibly small population (green) of protein is pre-activated. When the reaction is started by jumping the temperature and tipping the free energy profile a little, the tiny green population does not contribute much to the kinetics. Rather, the observed kinetic trace results from slow interconversion of the large orange populations (folded and unfolded) over the barrier, yielding an exponential decay signal with time constant τ_a

in Figire 1.2D. Now consider the lower free energy profile in Fig. 1.2C. Its low barrier supports a large pre-activated population of protein (green). The moment the temperature is jumped and the profile is tipped, these proteins react promptly in time $t < \tau_m$, giving rise to a very fast decay of the kinetic trace in Fig. 1.2D. The orange population has to be activated and kicks in much later, again folding with rate coefficient k_a. Thus the 'low barrier' kinetic trace in Fig. 1.2D has two phases, a fast one below $t = \tau_m$ (green), and a slow one (orange). It is as though the rate coefficient k is really large when $t = \tau_m$, and then drops to a constant value of k_a. This idea is illustrated in Fig. 1.2B.

Another way of thinking about it is as follows: there are always some proteins (green) that fold downhill when a folding reaction is initiated. When the barrier is large, those proteins are an invisibly small fraction of the total ensemble. However when the barrier is small, they make an easily detectable contribution to the signal. In terms of single molecule traces (Figure 1.2D), this line of reasoning is also evident: a fast folding protein spends a much larger fraction of the time making transitions (green, related to τ_m), whereas a slow-folding protein spends most of its time waiting between transitions (orange, related to τ_a).

The 'molecular time', the 'transit time' to diffuse across the transition region, and the 'speed limit' are related concepts, although defined differently. The 'speed limit' of protein folding is the fastest time it would take a protein of certain size to fold on a free energy landscape with the smallest possible barrier. For most of the fast-folding proteins the speed limit is estimated to be on the order of 1 μs. The 'transit time' would be about twice τ_m because a successful diffusion to the barrier takes about as long as the diffusion away from the barrier. Exactly how long it takes a protein to transit across the activated region of the free energy depends on the computational or experimental observable (reaction coordinate) chosen to monitor the crossing process (Fig. 1.2A). For this reason we will not distinguish τ_m, 'transit time' and 'speed limit' here. A variety of scalings have been proposed for the 'speed limit': a logarithmic scaling with absolute contact order (which increases with protein size and complexity by measuring the length scale of non-local contacts between amino acids)[11], and an inverse scaling with sequence length[59].

Since a protein cannot fold faster than the elementary steps that are re-

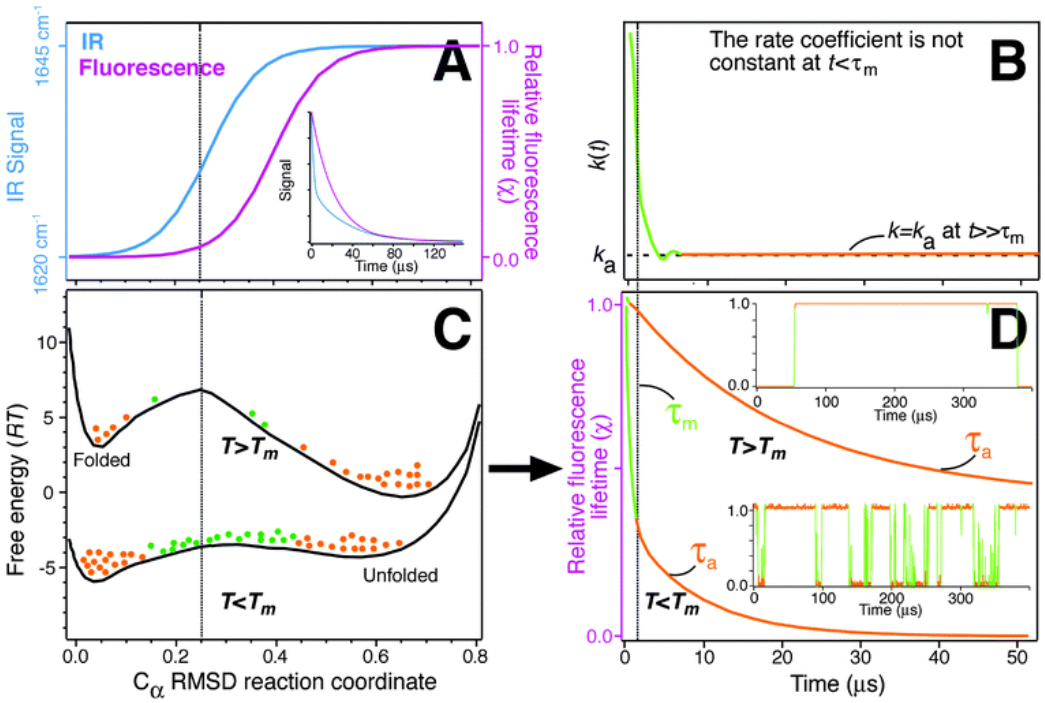


Figure 1.2: Folding from activated to downhill, using WW domain experiments and simulations as an example. (A and C) The bottom shows schematic free energy profiles[5, 6] along the C_α-RMSD reaction coordinate at temperatures below and above the melting temperature T_m. Protein populations in folded/unfolded minima (orange) and near the barrier (green) are shown as dots. The top (A) shows two probe signal profiles. They increase monotonically with reaction coordinate, and can therefore serve as experimental reaction coordinates. Different probes progress differently as the protein folds,[6] yielding probe-dependent kinetics particularly when the barrier is low and the protein population samples the transition region (vertical dashed line). (B) On the molecular time scale τ_m (vertical dashed line) the rate coefficient is not a 'rate constant', but depends on time[7]: pre-activated population (green in C) reacts promptly, much faster than population that needs to be activated (orange in C). Only later does the rate coefficient 'settle down'.

Figure 1.2: (Previous page.) (D) In the example, at $T > T_m$ the pre-activated population is negligible, and exponential-decay kinetics with a 'slow' time constant ?a is observed. At $T < T_m$, the pre-activated population is large, and a prompt phase precedes the 'slow' exponential-decay kinetics. The insets in (D) show corresponding single molecule traces: for a high barrier ($T > T_m$ in the example), the activated protein (green), is sampled rarely. For a low barrier ($T < T_m$), the activated protein is sampled frequently. In essence, there are always pre-activated proteins that fold promptly downhill at the 'speed limit.' If the barrier is large, this population is unobservably small due to the Boltzmann factor. If the barrier is small, the population becomes easy to observe.[8] The terms 'molecular time', 'speed limit', 'transition state transit time', 'downhill folding time' refer to the same time scale, but are not identical. Also, the decay in (D) at $t < \tau_m$ is not necessarily an exponential with time constant τ_m,[9] although frequently fitted as such.[6, 8, 10]

quired for structural assembly, much effort has been directed at identifying the speed of such elementary steps: loop formation in the unfolded state,[84–86] nucleation and growth of secondary structure,[51, 87–92] and internal friction that controls the dynamics of a collapsed polypeptide chain[93–95]. We consider an example of each in turn in the Support section.

1.2.3 Support

Loop formation in the unfolded state was one of the first folding events to bridge the gap between experiments and simulation. For example experiments on Cys-(Ala-Gly-Gln)$_k$-Trp peptides, where k ranged from 1 to 6, yielded the time constants for loop formation probed by Trp quenching by Cys ranging from 40 to 140 ns.[96] These experiments were simulated for $k = 1$ and 2[97] and the time constant for contact formation between Cys and Trp of ≈10 ns was found, faster than determined experimentally. Based on the simulations it was concluded that the rate of loop formation is reaction-controlled, not diffusion-limited.

Several recent studies of α-helix formation are especially noteworthy because of the remarkable time resolution that has been achieved experimentally and because of increased reliability of the modern force fields. To study secondary structure formation, Ma *et al.* developed an ultrafast temperature jump instrument.[98] They used this setup to measure ultrafast

dynamics of α-helix formation of a 5-residue peptide called W_1H_5, which consisted of a tryptophan and a histidine connected by three alanines, Ac-$W(A)_3H^+NH_2$.[99] They found that the dynamics of α-helix folding after a temperature jump involved two time scales: a fast time scale with a time constant of ≈450–850 ps and a slower process with a time constant of ≈3.6–5.3 ns with faster kinetics corresponding to the higher temperature for both time scales. The authors attributed the fast time scale to the annealing of the folded structure, while the slow time scale was assigned to the diffusion to a collapsed state. These results were later investigated computationally by De Sancho *et al.*[100] using replica exchange molecular dynamics simulations with AMBER ff03w force field. They found that their results were in good agreement with experiment because even though multiple phases could be extracted from the computational model, the dynamics could be approximated well with a double-exponential function. However, the authors came to a conclusion that the slow time scale observed in experiments was most likely due to the shrinking of the helical peptide at the C-terminus and interconversion within the collapsed ensemble was responsible for the fast time scale. Helix nucleation, which is defined in this work as the organization of three consecutive amino acids into a helical geometry, was still proposed to occur on the order of 20–70 ns. A similar time constant for helix nucleation is also supported by several other studies but it is important to note that helix formation time scales reported in the literature vary from tens of nanoseconds to several microseconds because of indirect probes and different model systems used.[88, 101, 102]

Although β-hairpins are generally thought to fold more slowly, very recent work has shown that they can also zip up from both ends on a sub-microsecond time scale.[103] Even β-hairpin peptides with more complex kinetics, such as the tryptophan-rich 12-mer called trpzip, have fastest experimental time scales on the order of a microsecond.[104] Snow *et al.*[105] did a comparison between simulations and temperature jump measurements of trpzip variants. They found reasonable agreement between experiment and simulation for 2 out of 3 variants that they studied. The third variant had a propensity for kinetic trapping, which indicated the need for more accurate potential energy functions.[106]

After a protein reaches a compact state, its motions become limited not only by diffusion through the solvent, but also by the internal interactions

of different parts of the polypeptide chain. These interactions slow down the folding process by constraining the torsional angles of the protein chain and preventing it from reaching the native state efficiently. The internal friction concept originates from polymer physics.[107] Schuler and colleagues have used single-molecule FRET and correlation analysis to quantify unfolded state dynamics and internal friction in unfolded proteins.[108, 109] They studied a small cold shock protein from *Thermotoga maritima* called Csp and found that the effects of internal friction are less significant at high denaturant concentration when the protein is expanded but become more pronounced at lower denaturant concentrations when the protein is collapsed. Additionally, they investigated internal friction in two intrinsically disordered proteins (IDP) and discovered that internal friction depends on the sequence of amino acids that constitute the protein and also on the charge repulsion between amino acids, which is an important consideration for IDPs. The small protein trp-cage also illustrates the application of internal friction ideas to experiment and simulation. Qiu *et al.*[110, 111] found a linear dependence of the observed rate constant after the temperature jump on solvent viscosity ηs in correspondence with Kramer's theory, $k_{obs}^{-1} = a + b\eta s$. A linear fit to the data does not extrapolate to zero but rather a ≈700 ns. They concluded that solvent viscosity controls protein folding when $\eta s > 100$ P^{-1} (viscosity of water at 293 K is 1 cP = 100 P^{-1} = 1 mPa · s), but below that value the process of folding is governed by other factors including intra-chain diffusion. They argued that the fact that it is the time scales and not the rates that are additive means that the two reaction control mechanisms are sequential. Zagrovic *et al.* did Trp-cage simulation in generalized Born/surface area (GB/SA) implicit solvent at different solvent viscosities (from as high as that of water to as low as 10^{-4} times that of water).[112] They found a $k_{obs}^{-1} \approx \eta s$ relationship at viscosities 10 P^{-1} $< \eta$s $<$ 100 P^{-1} and $k_{obs}^{-1} \approx \eta s^{1/5}$ power law for smaller viscosities. They concluded that low viscosity MD cannot be used to extrapolate the rates of protein folding using the normal Kramer's scaling of $k_{obs}^{-1} \approx \eta s$. The approaches in this example agree that viscosity must be corrected to account for the protein acting as its own solvent. They differ in what description is best; as discussed in Section 1.4, such differences can arise from the ambiguities inherent in coarse-graining folding from a high dimensional process to just one reaction coordinate.[113]

Chain rearrangement times and relaxation of unfolded states provided the

first indirect estimates of the 'speed limit' value of τ_m.[114, 115] The first direct measurement of τ_m was an ensemble T-jump experiment on lambda repressor fragment.[8] These experiments directly observed the theoretically predicted settling of the rate coefficient into a constant value at $t > \tau_m$, after which two-state folding could be described by a constant rate coefficient k_a and a single-exponential decay. It was shown by T-jump experiments for lambda repressor (α-helix bundle) and WW domain (a three-stranded β-sheet) that the 1 μs molecular phase accounts for more and more of the kinetic amplitude when the protein is stabilized. Simulations of Fip35 WW dynamics on a one-dimensional potential surface[6] derived from single-trajectory MD simulations[116] that observed many folding/unfolding events were fully consistent with the experimental results. It was also shown that the transit process is heterogeneous because a stretched exponential is required to fit the short time relaxation dynamics of proteins when high signal-to-noise is achieved in experiments.[9]

Initial single-molecule experiments were not successful at obtaining a firm value, but were able to show that $\tau_m < 250$ μs,[117, 118] an upper bound consistent with ensemble measurements.[6, 8] Single-molecule detection of the transit time for folding is an important but challenging task: important because single-molecule experiments can examine heterogeneity of the transition state ensemble in great detail, and challenging because a statistically significant sample of photons emitted from FRET probes must be collected while the protein is diffusing across the barrier. For RNA folding, Neupane *et al.*[119] used an optical trap do determine the upper limit of the transition path time for several structures. They arrived at the instrument-limited value of $\gg$50 μs. However, when they analyzed one-dimensional free energy landscapes that resulted from their experiments, they estimated the transition path times of $\approx$2–6 μs for most of their samples. In a very recent paper, Chung *et al.*[120] report measurements of the transition path times of two proteins, which differ in their folding rate coefficients by 4 orders of magnitude. The transition path times only differ by a factor of 5 for these proteins: 2 μs for FBP28 WW domain and an upper bound of 10 μs for GB1. The result for FBP28 WW domain is consistent with ensemble measurements[6] as well as simulations[116] of τ_m for the much faster-folding Fip35 WW domain.

1.3 Lowering the barrier to go downhill

1.3.1 Key concept

'Downhill' protein folding refers to scenarios where the protein folds without a significant free energy barrier (less than several RT) due to quasi-perfect cancellation between the enthalpy and entropy contributions to free energy along the whole reaction coordinate. Many examples of downhill protein folding have been seen experimentally and by molecular dynamics simulation.

1.3.2 Elaboration

One of the major confusions in the folding literature is between energy landscapes and free energy landscapes. The folding enthalpy $\Delta H(S, P)$ of globular proteins is generally negative above room temperature: folding becomes increasingly exothermic at higher temperature.[121] The folding entropy $\Delta S(E, V)$ is also negative at sufficiently high temperature: the polypeptide chain organizes during folding. For this reason, a plot of contact enthalpy vs. chain entropy slopes like a funnel as the protein folds (Figure 1.3C). The protein goes 'downhill' in the funnel-shaped enthalpy surface. Due to non-native contacts or 'traps', the funnel is rough and does not go 'downhill' completely smoothly. In that way the population picture (traps, native states, *etc.*) and the energy picture nicely connect.

'Downhill' in the enthalpy funnel is not at all the same as the downhill folding discussed in the literature, which relates to free energy landscapes, not energy landscapes. In the laboratory, measurements are usually made at constant temperature and pressure, so the Gibbs free energy of folding $\Delta G(P, T) = \Delta H - T\Delta S$ is the natural thermodynamic potential obtained from the enthalpy by Legendre transform[122]. The negative ΔH and positive $-T\Delta S$ tend to cancel, leading to very small free energies for folding, on the order of 0.5 kJ mol^{-1} per residue (Figure 1.3). Entropy favors unfolded polypeptide conformations, enthalpy favors folded ones, leading to free energy minima for the unfolded and folded states. In-between, cancellation is imperfect, producing free energy barriers and intermediate states such as for example traps stabilized by non-native contacts. However, the enthalpy–entropy cancellation is quite good, which is why protein folding reactions are

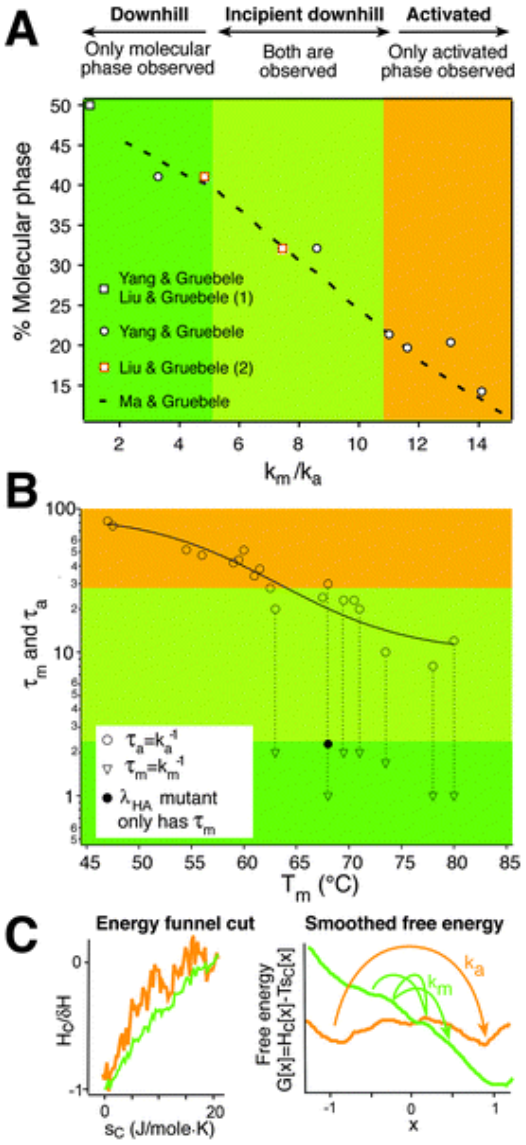


Figure 1.3: Experimental signatures of downhill folding upon protein stabilization: only the fastest-folding, most stable mutants of lambda repressor fragment have a significant population undergoing prompt reaction (the molecular phase shown in Figure 1.2D). (A) The measured molecular phase amplitude increases smoothly when the activated rate k_a increases towards the molecular rate $k_m \approx 1\mu s^{-1}$ (Yang & Gruebele[11]; Liu & Gruebele (1) and (2)[11, 12]; Ma & Gruebele[9]), as predicted when the free energy barrier approaches RT (downhill folding).[13]

Figure 1.3: (Previous page.) (B) The kinetics of mutants that are relatively unstable can be fitted by slow single-exponential kinetics upon temperature jumps (orange area); their activation barrier is too high to carry a measurable pre-activated population. Mutants that have $T_m > 60$ °C show an additional fast molecular phase (triangles) because their barrier is low enough so there is a promptly reacting (downhill folding) protein population (C) On the left: the normalized enthalpy of the polypeptide chain generally decreases when the configurational entropy s_c decreases: as favorable contacts are made, the polypeptide chain moves less freely. Folding is 'downhill' in enthalpy (folding is an exothermic reaction), resulting in an 'enthalpy funnel', but this is not what is meant by downhill folding. On the right: the free energy G can be computed from the enthalpy and entropy as a function of an arbitrarily chosen reaction coordinate x by evaluating H and S at x and averaging over all other orthogonal coordinates. 'x' could be the radius of gyration, distance between two FRET labels, *etc.*, and is normalized from -1 (unfolded) to 1 (native) here. Of course, a carefully chosen set of coordinates $x, y, ...$ provides a more complete description of a reaction as complicated as folding than just a single coordinate x. The free energy has a barrier (orange) when the enthalpy does not funnel the protein towards the native state efficiently enough to offset the decreasing entropy (orange funnel on the left). The free energy is downhill (green) when the exothermicity of the reaction is sufficient to offset the loss of entropy everywhere along the reaction coordinate (green funnel on the left). The protein then folds with the molecular rate k_m instead of the slower rate k_a (black circle in (B)). In intermediate cases both rates can be measured simultaneously (triangles + circles in (B) connected by a dot, or $T < T_m$ trace in Figure 1.2D), allowing an absolute determination of the free energy barrier height[6, 8–12].

so fast at room temperature compared to many other chemical reactions.

When the cancellation of ΔH and ΔS is further optimized (by protein engineering or natural evolution), even the free energy goes downhill. That is downhill folding (Figure 1.3C). We define reactions with barriers <3 RT (about 7.5 kJ mol^{-1}) as downhill or incipient downhill folding reactions. The choice of "3" is somewhat arbitrary, and corresponds to a $P \approx 5\%$ Boltzmann population at the barrier top, detectable by an experiment with a signal to noise of about 20:1. In other words, the fundamental assumption of transition state theory that the barrier-top population is negligible, has been violated by 5%. At 1 RT, the violation is over 25%. ($P = \frac{e^{-\Delta G/RT}}{\Delta}$, where Δ is the partition function.)

The discovery of natural and engineered microsecond folders is half of the equation that allows a direct comparison of experiment with simulations on the microsecond time scale. The other half is the improved force fields[14, 21, 22] and faster computing that allow not just the barrier crossing itself, but the slower kinetics (the waiting for the barrier to be crossed) to be simulated. Improved force fields are of key importance here: computational power is useless if a protein folds into the wrong state, or even unfolds from its native state, as was the case with early force fields.

1.3.3 Support

We begin our comparison with a 'long single-trajectory study'. Lindorff-Larsen *et al.*[5] simulated 11 small proteins (and the hairpin chignolin) previously studied by fast protein folding experiments. The largest of these proteins, the 80-residue lambda repressor fragment, approaches the average size of globular protein domains (≈120 residues). They observed at least 10 folding and unfolding events for each protein. For these 11 proteins much secondary structure formed before the longer-range native contacts. They found that across the protein set, the unfolded state contained residual secondary structure (16% α-helices and 5% β-sheets). Residual unfolded structure is an important feature that promotes fast folding (Section 1.5).

For 9 out of 11 simulated proteins, folding events could be clustered into 2 to 3 folding pathways. These pathways shared 60% of native contacts on average, so to that level of accuracy, each protein folded on 'a pathway. For the two remaining proteins (NTL9 and G), several distinct folding pathways were identified based on the order of β-sheet formation. Even proteins that preferentially formed one β-turn first (WW domain)[116, 123] still have minority populations (10–20%) forming the other turn initially. These results again highlight the importance of energy *vs.* population (Boltzmann factor). The ensemble of low-energy folding pathways was heterogeneous, but with sufficient coarse-graining most of the population preferred one or two paths. As another example, the lambda repressor fragment simulated by Lindorff-Larsen *et al.* has been shown experimentally to fold with or without highly helical intermediates, and with or without rapid collapse to a small radius of gyration, depending on mutations and solvent condition.[2, 79] The same

final fold formed in rather different ways, but each specific mutant or solvent condition had a preferred path. As discussed earlier, the presence of multiple paths on the energy landscape may have evolved to confer robustness to the folding process, even if a specific path carries most of the population under specific circumstances.

All 11 proteins simulated by Lindorff-Larsen *et al.* had a compact and native-like transition state ensemble with a folding barrier <4.5 RT. BBL, protein B, and homeodomain even folded without an identifiable barrier. This result highlights the near-perfect cancellation of enthalpy and entropy that leads to low-lying paths on the free energy landscape (but see the caveats below). For two of the computed proteins, absolute barrier heights have been determined from experiment by comparing the molecular time τ_m required to equilibrate barrier-top population with the time τ_a required for the activated reaction to occur: $\Delta G_a/RT = ln(\tau_a/\tau_m)$[8]. For lambda repressor, experiments showed that the amplitude of the molecular phase increases when the protein is stabilized,[8] in accord with kinetic theory[7]. For the D14A/Y22W/G46,48A mutant, a barrier of $\Delta G^\dagger = 1.5\ RT$ was determined, exactly the value later extracted from molecular dynamics[5] using the method developed by Hummer[124]. Microfluidic mixer experiments by DeCamp *et al.*[125], where the bias towards the native state was stronger than in T-jump experiments, observed complete downhill folding of a lambda repressor fragment. For the WW domain variant FiP35, experiments by Liu *et al.* measured an absolute barrier height of 3.3 RT[10], whereas the folding/unfolding trajectories simulated an unfolding free energy of $\Delta G^\dagger = 3.5\ RT$[116]. Reasonable to excellent agreement is also obtained in other cases. Thermodynamic studies of a protein BBL mutant found no barrier[126], in perfect agreement with the coarse-grained free energy surface computed by Lindorf-Larssen *et al.*, Qiu *et al.* used laser temperature jump to determine that the folding time constant of trp-cage is 4.1 μs, which corresponds to ≈4 RT barrier[127]. The folding time constant observed for trp-cage in simulations was 14 ± 4 μs[5]. Zhu *et al.*[23] reported microsecond folding of α_3D, a de novo designed protein with folding time of 3.2 ± 1.2 μs at ≈323 K. The folding time reported by Lindorff-Larsen *et al.*[5] was 27 ± 8 μs.

Although the discrepancies between molecular dynamics simulations and experiments could arise due to imperfections of force fields or differences in

conditions (solvent, temperature), these discrepancies could also result from the inability to describe experiment and modeling with a common reaction coordinate (Fig. 1.2A). Hence an important caveat: the time scale τ_m depends on the probe signal (reaction coordinate) used. A one-dimensional picture does not capture the full complexity of the folding process at the experimental and computational level of resolution now possible (Section 1.5). It should also be noted that much progress is still required to make simulations more accurate. For example, the computed melting temperatures of proteins are generally too high. Nonetheless, it is quite remarkable that a single force field could fold 11 proteins in quantitative agreement with some experimental results, indicating significant progress in MD force field development.

Designed peptides have been important testing grounds for simulations, and recently simulations have made predictions for re-designing fast folders. The first direct comparison between experimental and computed folding rates and equilibrium constants was for the peptide BBAW[67], a design based on BBA5 by Imperiali and coworkers[128] that added a tryptophan probe. Here the simulations were of the 'multiple shorter trajectories' type. 700 μs of total simulation time[67] yielded a small number of folding events, in agreement with an experimental T-jump folding time of *ca.* 1.7 μs. More recently, Bunagan *et al.*[129] used biased Monte Carlo replica exchange methods (BMCREM) to design a mutant of the 20-residue trp-cage (P12W) called Trp^2-cage, which folded in 1 μs and was $\approx$15 K more stable than the parent protein. Piana *et al.*[123] used a single multiple folding/unfolding trajectory to design the GTT mutant of Fip35 WW domain, which folded about three times faster than the original FiP35, itself a designed construct[10, 130–132]:$\tau_{obs} = 3.7 \pm 0.4$ μs *vs.* $\tau_{obs} = 13 \pm 4$ μs. The increase in folding speed was attributed to the preorganization of the unfolded state, highlighting the importance of residual unfolded structure (Section 1.5).

We have proposed that replacing charged functional residues by more hydrophobic residues, and longer functional loops by shorter loops, will redesign a protein towards downhill folding[10, 11, 130–134]. This design principle of function frustrating efficient folding has been used to design faster-folding WW domains[130, 135] and lambda repressors[133]. The same idea has also been put forward by Gai and coworkers: five mutations of an albumin binding domain were predicted to increase the hydrophobicity of the

protein in a favorable way to increase the folding speed[136]. They found a linear relationship between the *log* of maximal folding rate and mean hydrophobicity. Wang *et al.* showed that a K5I/K39V mutant of this protein has a folding time constant of only 1.2 μs[137]. They argued that this protein approaches the speed limit because of its highly optimized hydrophobicity.

There is now a rich interplay between experiments and simulations[64, 123, 138, 139]. Consider the following model system: the C-terminal domain of chicken villin headpiece. Depending on the variant, this protein contains 35 (HP-35) or 36 (HP-36) amino acids, which arrange into a three-helix bundle in the native state. HP-36 holds the distinction that Duan *et al.*[46] ran the first all-atom explicit folding simulation longer than 1 μs. Freddolino *et al.*[140] subsequently identified a trapped state. On the experimental side, Kubelka *et al.*[141] introduced a F35A mutation to test Pande's prediction[142] that mutating out F35 would increase the folding rate by eliminating the off-pathway intermediate in which F35 docks into the hydrophobic core. Buscaglia *et al.*[84] used quenching of the triplet state of tryptophan to study the dynamics of C-HP-35. "C" stands for cystein that was introduced into the protein as a quenching probe. The results showed two phases, consistent with previously reported T-jump data. Piana *et al.*[143] very recently did an autocorrelation analysis of folding of HP mutants and found – in good correspondence with experiments – two phases with time constants of $\approx$100 ns and $\approx$5 μs. They also estimated the pre-exponential factor to be $\tau_m \approx 0.5$ μs to 1.5 μs, in good agreement with 'speed limits measured for lambda repressor and WW domain.[6, 8]

As computing power has improved to reach experimental downhill folding timescales, so force fields have improved to reach experimental accuracy. Villin headpiece subdomain has recently been used to compare four molecular dynamics force fields[22]. The experimental folding rate was reproduced well by all four force fields, and all force fields folded villin into a native-like state. However, the flux through different folding pathways depended on the force field, indicating that the unfolded state and the pathway are not as robust as the folded state (Figure 1.4). Thus force fields are good at describing the lowest energy state, but higher lying states, important for protein functions and for denatured protein structure, are not as well described. This problem exists partly because the current goal is for force fields to produce native-like states; higher energy states will become the next grand goal, opening the

door for better quality simulation of function (as opposed to just the native structure).

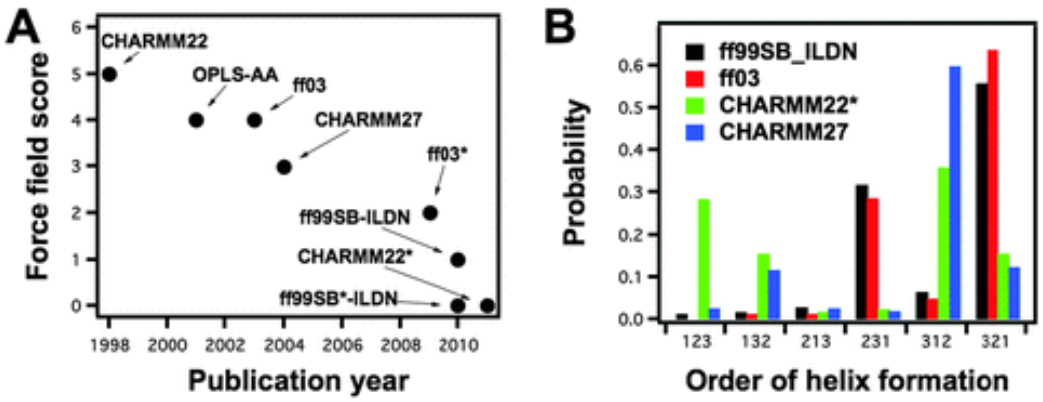


Figure 1.4: Systematic benchmark studies of empirical force fields to improve their performance in folding simulations are now more frequently appearing in the literature[14–20]. (A) Various force fields were tested by Lindorff-Larsen *et al.*[21] and a score based on the performance of the force field against the chosen model systems was devised such that the lower score indicates better agreement with experimental data. The plot shows the improvement of force fields over time. (B) Piana *et al.*[22] used four different force fields to fold villin headpiece. Although all simulations arrived at the correct native state, the folding mechanism depended on the force field used. The panel shows that the flux through different reaction pathways (123 *etc.* is the order in which the three helices of villin headpiece form) is a function of the force field that was used.

The other main challenge in molecular dynamics simulations has been the insufficient sampling time. The problem of sampling has been overcome to some extent, as nowadays resources exist to simulate a protein for up to milliseconds in atomistic solvent. Long simulations allow researchers to identify inaccuracies in force field parameterization and sometimes even fix these issues. For example, Freddolino *et al.*[14] simulated Pin1 WW domain, a three-stranded β-sheet, using the CHARMM22 force field with CMAP corrections. In their simulation the protein formed α-helical structures instead of the crystallographic native state. They used the deactivated morphing methodology to determine whether these structures were kinetic traps or thermodynamic inaccuracies in the force field[144]. They came to the conclusion that the force field parameterization of linear *vs.* bent hydrogen bonds was imperfect and favored α-helices over β-sheets by 15–30 kJ mol^{-1}.

Mittal *et al.*[145] also used a force field that was biased towards α-helices,

AMBER ff03, to fold villin headpiece HP-35, a 3-helix bundle. Then they introduced a correction to the backbone potential within the force field (making it AMBER ff03*) and managed to fold Pin WW domain with it, an all-β-sheet protein, which was the first time for a β-sheet protein to be folded with a version of AMBER ff03 force field. Lindorff-Larsen *et al.*[21] used extensive single-trajectory simulations to systematically analyze the force field quality. They did simulations on folded proteins, peptides with certain secondary structure propensities and two proteins that have α-helical and β-sheet native states, villin and GTT Fip35 WW domain. They concluded that force fields are getting better with time and that at present AMBER ff99SB-ILDN and CHARMM22* force fields reproduce experimental results better than others (Figure 1.4).

Although currently the results of molecular dynamics simulations still depend on force fields and no one perfect force field that the community has agreed on exists, these discrepancies also highlight that energy differences between competing folding pathways are small: although one force field or another may put the wrong path lowest in free energy, it in fact is one of the lowest free energy paths in reality. Certainly examples are known where one protein sequence can fold into a more β-rich or α-rich functional state[146]. Thus even nature flips energies of pathways when solvent conditions or temperature change.

1.4 How many reaction coordinates do we need?

1.4.1 Key concept

The number of reaction coordinates that will be sufficient to understand the folding of a protein depends on the level of coarse-graining that one is willing to tolerate. Better reconciliation of experimentally and computationally accessible reaction coordinates is an important goal for the near future.

1.4.2 Elaboration

A question that dogs the field of protein folding in general, and the direct comparison of microsecond experiments and simulations in particular, is the

nature and number of reaction coordinates best suited to describe the folding process. Early simulations showed that one coordinate cannot provide a full description even for small peptides:[147] there may be a predominant pathway, but it is only that – predominant. Likewise even some millisecond or slower experiments demonstrated parallel paths,[3] a clear sign of multiple reaction coordinates.

How many coordinates? This has been a contentious question, but the answer for a complex process like folding is clear: the number depends on the level of coarse-graining at which we want to understand the process. At one extreme, every backbone and side-chain torsion angle (we can safely ignore stretching and bending motions to a first approximation) is a coordinate. At the other extreme, one predominant coordinate is singled out, leading to equation 1.1. For example an experiment may show with 20:1 signal to noise ratio that some secondary structure forms first, followed by collapse to a small radius of gyration, followed by burial of tryptophan, in that order. We can draw a single coordinate axis, and even put quantitative numbers on it[116]. The one-dimensional view of protein folding is currently very common in the literature. The coarse-grained picture is very useful because Kramer's theory does apply reasonably well to many proteins at a low level of resolution (see Section 1.2). Nonetheless, both single[5, 116, 148, 149] and multi-trajectory[67, 142, 150, 151] simulations have shown that the overall folding process is built up from many faster interconversions on a multidimensional energy landscape.

What coordinates? This difficult question is currently the greatest divide between experiment and simulation. "Fraction of native contacts" is a perfectly acceptable reaction coordinate that can be computed easily. However, no experiment will ever measure it by its literal definition (residue pairs with direct side chain contact in the native state that approached within an arbitrarily chosen cutoff distance, and are assigned a weight of 0 or 1 based thereon). "Circular dichroism at 222 nm" is a perfectly acceptable reaction coordinate for average helix content in an all-α-helical protein that can be measured easily. However, not even quantum-based models are likely to compute it accurately any time soon. Some progress has been made to cross-validate coordinates between fast folding experiments and simulations, including fluorescence *vs.* fluorophore solvent-exposed area[63, 123], or two-dimensional infrared spectra and β-sheet content[152, 153]. This process

needs to continue so we can: (1) quantitatively compare computed and measured reaction coordinates to provide a satisfactory description of folding; (2) determine how linearly independent different coordinates are from one another, so the most informative sets are computed or measured; (3) understand functional excursions of the folded state towards less/differently folded states in terms of rigorous coordinate sets to better describe higher energy functional states.

What about topology? It has been clear for some time that it would be desirable to describe folded structure in general terms rather than *via* atom-by-atom numerical coordinates[154–156]. Many folds are intuitively related, even if they differ in quantitative detail, and structural classes and families have been identified[157]. In folding science, the term topology has been used in different ways, not necessarily making a rigorous connection with its mathematical meaning. A useful set of criteria for distinguishing and identifying topologies has been proposed in terms of Gauss integrals[158], and has been applied to fast folding simulations[5]. Gauss integrals are ideally suited to describing how a worm-like chain intertwines with itself. Other parameters such as absolute contact order have also been useful in merging protein size and fold complexity into a single number[159].

Microsecond folding experiments and simulations have shed light on the aforementioned issues. The use of multiple probes in experiments[9, 13, 24, 104], as well as hidden Markov analysis of multiple trajectories[70, 160–164], or single long trajectories[5] all have shown that multiple low energy paths exist, although many of them are not highly populated due to their exponentially decreasing Boltzmann weight. (see the Support section for more information)

An arsenal of new experimental methods, including two-dimensional infrared spectroscopy[152, 165–168] and resonance Raman[169–172] is now coming online. At the same time, molecular dynamics simulation provides a rich set of coordinate information at various levels of coarse-graining[160, 161, 164]. The problem is that atomic coordinates are not sufficient to compute accurately experimental reaction coordinates such as circular dichroism at 222 nanometers (a stand-in for α-helix content). Other experimental reaction coordinates fare better, such as the radius of gyration, which can be computed accurately from simulations, including even a solvent correction[173]. Yet others are in-between, such as a pairwise distance between FRET labels.

The actual FRET-labeled construct is not usually simulated, but distance between the residues where labels are connected, coupled with assumptions about random rotational orientation, yield a proxy to experimental FRET. Only when multiple measured and simulated coordinates can be compared quantitatively, will folding science be able to take the next step towards a multi-dimensional description of folding. An important goal will be to establish fully quantitative correlations between measured coordinates and their computed stand-ins. For example, tryptophan wavelength shift (experimental) and tryptophan solvent exposed area (computed) will have to be compared over a wide range of different folded and unfolded proteins. Much experimental data is already available. Doing the many needed simulations is no longer outlandish with increased computational power and better force fields. Likewise, better stand-ins can be computed thanks to vastly greater computational power. For example, when many microseconds of trajectory for multiple fast folders are becoming practical, one could simulate the actual FRET construct and do orientational averaging over the dipole–dipole coupling. Another possibility would be to simulate many different protein mutants for rigorous comparison with mutation experiments. Such simulations are already coming out, for example for different WW domains[123], different lambda repressor mutants[63, 149], and different versions of villin headpiece[140]. These results will settle many of the debates about the appropriateness of the Kramers' equation, or how well coordinate-dependent diffusion coefficients complement one-dimensional models to maximize their descriptive power of the folding process.

1.4.3 Support

Ma *et al.*[9] used two simultaneous probes, tryptophan fluorescence lifetime and the infrared amide I′ band to measure the relaxation of a lambda repressor mutant after a temperature jump. Fluorescence ($\lambda \approx 350$ nm, $\tau_{fl} \approx 3$ ns) probes protein collapse around a single tryptophan residue, IR ($\nu \approx 1650$ cm^{-1}), probes overall helix *vs.* random coil content, providing two very different reaction coordinates. IR and fluorescence yielded different kinetics 9 K below the unfolding temperature T_m, but converged to identical kinetics at T_m. Ma *et al.* concluded that near T_m, lambda repressor fragment

folded over a barrier: although IR and fluorescence are different experimental reaction coordinates, both switch from 'native' to 'denatured' on top of the barrier where population is small, and so both appear to change together. They concluded that 9 K below T_m, the free energy landscape switched to a downhill surface: without a barrier, the protein population passes at different times through the region where the IR and fluorescence probe switch. Ma *et al.* simulated these observations quantitatively with two-dimensional Langevin dynamics along the two experimental reaction coordinates. Working on the same protein, Dumont *et al.*[2] measured the radius of gyration (R_g) or 'compactness' of lambda repressor using small-angle X-ray scattering, and secondary structure using circular dichroism in stabilizing solvent (45% ethylene glycol in water) at T = 245 K. They observed excessive formation of helical secondary structure before collapse for some mutants, concomitant collapse and secondary structure formation for others. Long trajectory simulations have indeed revealed a variety of folding mechanisms for different lambda repressor fragment mutants[5, 148]. Again these results highlight that a predominant path usually exists for a mutant–solvent combination, but not for a specific fold topology.

Liu *et al.*[24] investigated folding kinetics of the designed protein α_3D[23] also by IR and fluorescence T-jump experiments (Figure 1.5). The observed rate was nearly temperature-independent by IR, but increased with temperature when probed by fluorescence. They could not reproduce the experimental results with a reasonable diffusion coefficient by Langevin dynamics along just one reaction coordinate. A two-dimensional description yielded a much more reasonable diffusion coefficient. The result of Langevin dynamics simulations depends on the reaction coordinate used: different probes switch at different times during the folding process, and the diffusion coefficient $D(x)$ is a function of position along the reaction coordinate x in coarse-grained pictures[11, 174]. For the latter case, Best *et al.*[174] found that $D(x)$ varies significantly along a reaction coordinate that represents fluctuations, but is mostly invariant when a reaction coordinate like fraction of native contacts is used. Full MD simulations show that the folding barrier of α_3D below the melting temperature Tm is less than 2 RT along a C_α-RMSD reaction coordinate, but the simulations have not yet been analyzed in terms of reaction coordinates closely related to the experimentally measured ones (*e.g.* helical content by IR, tryptophan quenching by fluorescence lifetime).

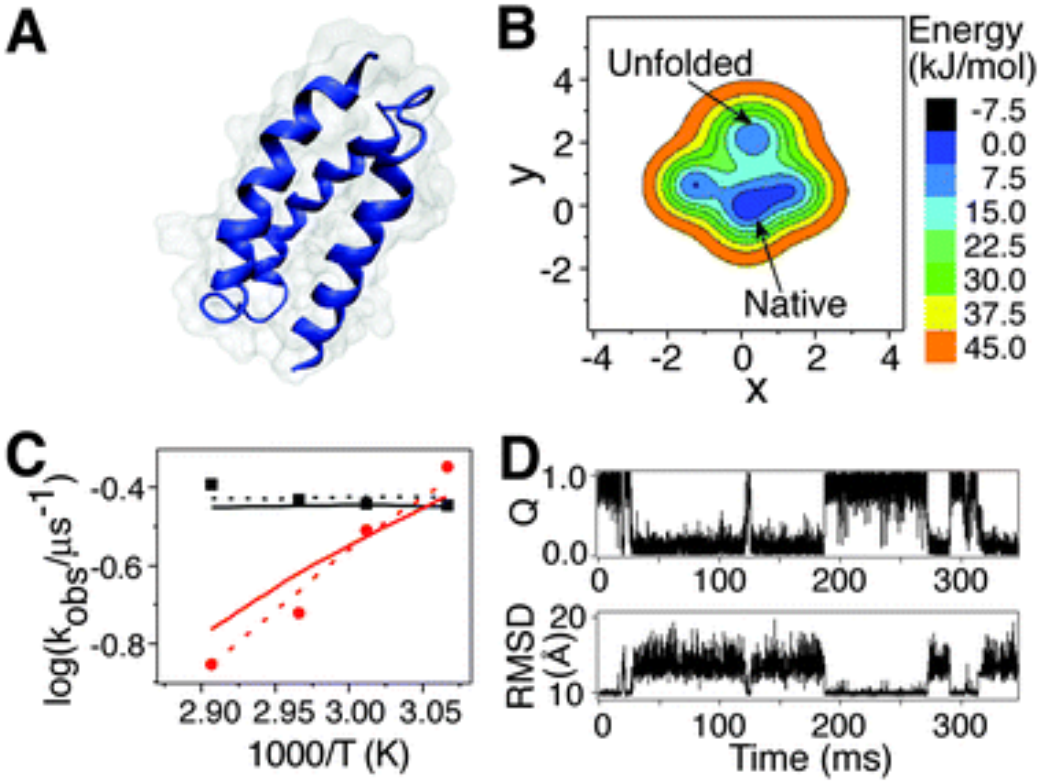


Figure 1.5: α_3D is a designed protein[23] for which a one-dimensional reaction coordinate cannot explain the probe-dependent kinetics using a reasonable diffusion coefficient[24]. (A) Structure of α_3D from PDB (2A3D). (B) A contour plot of the optimal two-dimensional free energy surface for α_3D calculated using Langevin dynamics as described by Scott *et al.*[25] (C) Kinetic rates measured with infrared absorption spectroscopy are shown as black dots, fluorescence spectroscopy as red dots. The dashed (one-dimensional model) and solid (two-dimensional model) lines represent the fits of the data using Langevin dynamics, but the one-dimensional fit requires an unrealistic diffusion coefficient; diffusion coefficients of incorrect magnitude or with unusual coordinate dependence are a warning sign that the model underestimates the dimensionality of the dynamics. (D) Time traces of Q and C_α-RMSD for α_3D from Lindorff-Larsen *et al.* show strong correlation, but are not equivalent[5].Q is the fraction of long-range native contacts. For the quantitative definition of Q see page 3 of Supplemental Materials in ref. [5].

1.5 Unfolded, trapped, misfolded

1.5.1 Key concept

Proteins did not evolve for stability or folding speed. They evolved to execute particular biological functions. Therefore, intermediates, partially unfolded proteins, and trapped non-native conformations that slow down the folding process should come as no surprise.

1.5.2 Elaboration

Proteins are evolved biological objects. They are not perfect at folding, but comprise a series of compromises: many proteins have to fold to function, so their polypeptide chain must be tightly packed to confer stability. Yet with only a 20 amino acid alphabet, positioning of functional amino acids is not perfect[175], so the polypeptide chain must be flexible to reach different functional states. Some proteins are disordered so they can bind/fold better.[176] Examples of proteins that exhibit functionality in the disordered state also exist.[177] It remains to be seen how disordered structures behave in the interior of living cells, where some of these proteins could be folded. For example, α-synuclein, a poster child for partly disordered proteins, is debated to be a well-folded tetramer *in vivo*.[178]

Folding stability and speed can be at odds with protein function, and microsecond experiments as well as fast simulation studies have quantitatively illustrated this conflict.[135] So it is natural for proteins to park 'needlessly' in traps from which they have to unfold before they can attempt folding again,[27] or to park in intermediates that slow down folding. 'Needless' refers to the folding process only, not to the function. In some cases, the states that are structurally intermediate between an extensively unfolded coil and the native state can be populated as part of protein function, or at least facilitate progress along a functional reaction coordinate. In other cases, difficult-to-fold parts of a protein could be the functional part.[135] In yet other cases, the trap may be a consequence of physical chemistry: for example, local β-sheet structure forms rather easily in denatured proteins,[179] and may simply be unavoidable even in the best evolved or engineered monomeric protein, as indicated by very recent studies.[1, 63]

Another important conundrum of protein folding is particularly well illustrated by fast folders and simulations on them: residual unfolded state structure. It is no secret that denatured proteins are not random coils, but contain residual short range order (*e.g.* α-carbon dihedral angle distributions with native-like averages[180]) and long range order.[181] Such residual structure can profoundly affect a protein's propensity to misfold, or influence its folding speed. For example, many downhill folders probably have relatively compact denatured states with native-like residual secondary structure; mutants with denatured states closer to random coils may fold much more slowly. Experimentally, it is very hard to characterize denatured states or intrinsically disordered proteins because a wide distribution of structures needs to be quantified. Computationally these states are equally problematic because of sampling issues. Since it is much more difficult to characterize the unfolded state structurally, most of the molecular dynamics simulations that are intended to fold proteins assume the absence of residual secondary structure as the initial condition. In the absence of structural information about the unfolded state this assumption is not totally unreasonable: proteins are synthesized vectorially on the ribosome and at least the N-terminus of proteins tends to be disordered some of the time. On the other hand, many proteins unfold and then refold many times over during their life cycle in the cytoplasm, but these unfolded states are likely to contain much residual structure.

1.5.3 Support

In some cases, a simulation has been able to suggest an initial structural ensemble for fast folding experiments. For example, Ensign *et al.*[150] did simulations on the double-norleucine mutant of villin HP-35 in which the relaxation rate and the number of observed kinetic time scales depended on the starting structure. The simulations that corresponded most closely to the experimental results[182] were initiated from a partially folded state (Figure 1.6), which the authors suggested to be a good approximation of an experimentally relevant unfolded state.

Some experiments and simulations indicate that non-native states can have rather slow dynamics, as opposed to the homogeneous nature of a random

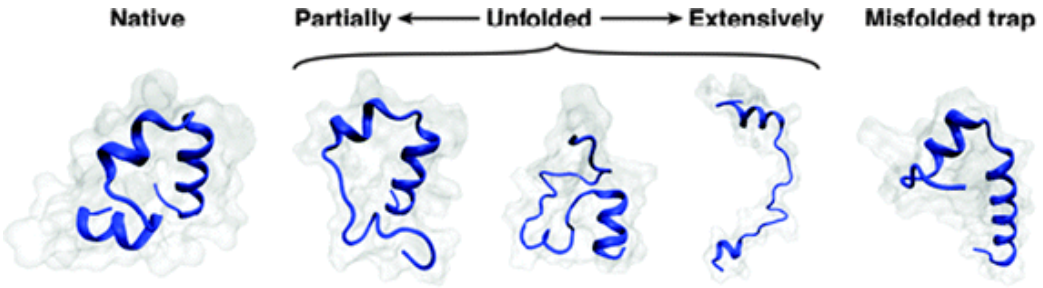


Figure 1.6: Various conformations of villin headpiece drawn from a simulation by Beauchamp *et al.*[26] The left-most conformation is the native state. Three structures in the middle broadly represent the unfolded state ranging from the partially disordered conformations that still resemble the native state to the significantly extended conformations with low residual secondary content. The right-most structure is a misfolded trap. Such traps lie off the predominant folding pathway.[27]

coil. For example, Waldauer *et al.*[183] used a microfluidic mixer to measure the intramolecular contact formation in the unfolded protein L and discovered that the diffusion coefficient in the absence of denaturant was very low. For this protein they proposed an upper bound for the folding speed limit of 20 μs, a much longer time scale than had been suggested previously.[59] Voelz et al.[28] made use of state-of-the-art computing methodologies by integrating the Folding@Home distributed computing system and calculations using graphical processing units to study the dynamics of the unfolded state of protein L (Figure 1.7). They also compared the results of their simulations with Trp-Cys quenching experiments and applied polymer theory to rationalize their findings. These results agreed with experiment that intramolecular diffusion in the unfolded state of protein L is much slower than expected for a random coil and that point mutations had a significant impact on the unfolded state ensemble. These and other such findings[184] suggest structural complexity of the unfolded state and the importance of non-native structure in the folding process.

Non-native traps may need some time to escape to more extensively unfolded states, which then fold rapidly to the native state. A specific example of such non-native traps was provided by Bowman *et al.*[63] They discovered a slow (10 millisecond) time scale in the folding of a mutant of lambda repressor fragment, in addition to the known fast kinetics. In the simulations, this slow time scale originated from non-native β-sheet-rich traps. Prigozhin

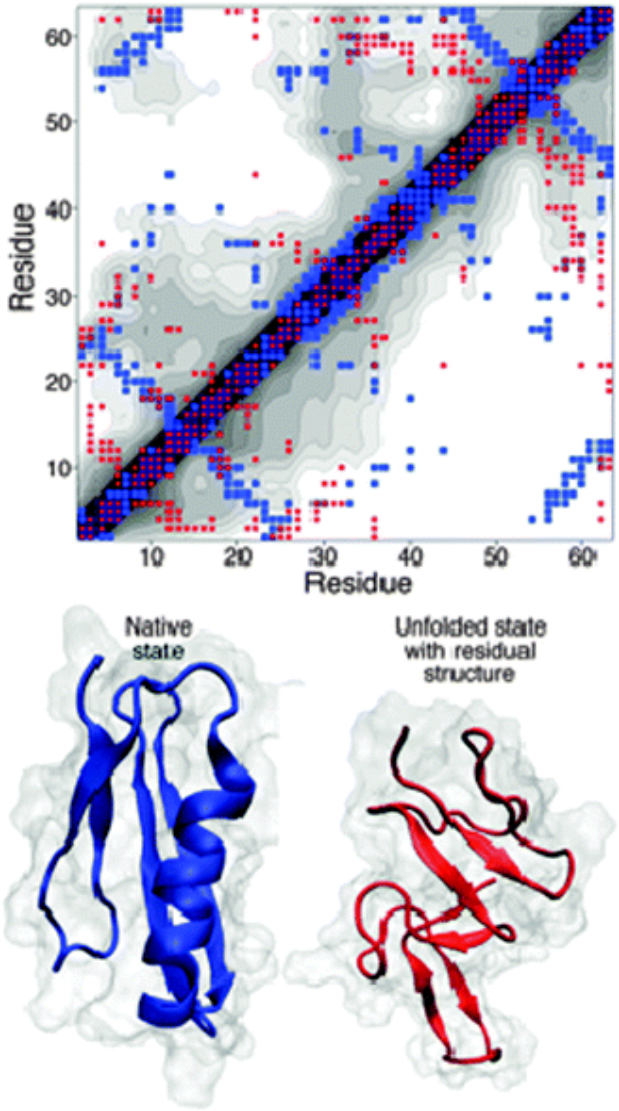


Figure 1.7: At the top: contact map of protein L.[28] The blue squares indicate residues that are in contact in the native state. The structure of the native state (PDB: 2PTL) is shown at the bottom-left of the figure. Data for the contact map were generated using the CMA server at http://ligin.weizmann.ac.il/cma/). The grey scale contour map shows the average distances (0–2 nm) between all pairs of residues at 300 K as simulated by Voelz *et al.*[28] The red dots indicate the contact map for one of the conformations of the unfolded state from the simulation. The structure of this state is shown at the bottom-right of the figure. Clearly this unfolded state is not a random coil but a partially structured non-native conformation.

et al.[1] observed a slow kinetic phase experimentally not with this mutant, but with a mutant differing only in one residue. It still remains to be discovered whether the experimental slow phase is due to β-sheet rich traps, as suggested by the simulations. But sequence-specific non-native structure clearly plays a role in this case also. If the slow phase in lambda repressor fragment does originate from compact β-sheet-rich structures, then the most likely explanation will be that the actual folding process (the interconversion between the native state and an extensively unfolded state) is fast, but getting out of compact off-pathway 'intramolecular amyloid' traps is a slow process.

Returning one last time to the importance of pre-organized structure for folding, Piana *et al.* recently were able to predict an interaction in FiP35 WW domain that slows down the folding process.[123] They analyzed long single trajectories with multiple folding/unfolding events and hypothesized that strand 3 of hairpin 2 of Fip35 WW domain, which does not make significant contacts with strand 2 of hairpin 1 in the transition state ensemble, could be engineered to form a more extended structure and dock against hairpin 1 early in the folding process thus stabilizing the transition state and accelerating the folding process. They proposed three mutations within strand 3 based on the Ramachandran angles of the amino acids in Fip35 WW domain. The new protein was called GTT by the names of three amino acids that were mutated into the Fip35 variant. Simulations of the mutant showed that it folded approximately two times faster than Fip35 WW domain. These results were validated by temperature jump experiments[123] showing that the observed relaxation time for the GTT mutant was $\approx$3 times faster than that of Fip35. The lessons we will learn from protein engineering and design driven by simulations will likely be instrumental in further improvement of force fields and will yield new model systems to continue the dialog between fast protein folding experiments and simulations.

1.6 Challenges met and challenges to come

Experiments and simulations of protein folding have come together on the microsecond time scale, helping protein scientists to understand folding in increasing detail. Even fast-folding proteins can visit a complex network

of low free energy states,[150] but a predominant path often exists for a specific sequence in a specific solvation environment.[5] Populations always look less heterogeneous than the underlying free energy landscape, thanks to the exponential Boltzmann factor relating population and energy.

As the next step, even more important than extending simulations to larger proteins and longer time scales, the time is ripe for a more rigorous comparison of simulation with experiments using consistent sets of reaction coordinates. Such comparisons will have to focus on more than one coordinate, so experiments can provide strong mechanistic constraints on simulations. Current force fields are approaching a level of quality where native states of small proteins can be obtained by direct physical modeling of the folding process, but not the specific mechanism by which they are reached. Correct mechanism involves the proper sorting of low-lying paths and states, highlighting again their important presence on the free energy landscape.

Although low lying states may be a nuisance in simulations, they are likely to play important roles. They may pre-pattern the functional dynamics of a protein as it explores higher energy states during function. Conformational selection or induced fit of enzymes would be a good example. Alternatively, such states may provide routes for protein evolution, which could switch alternate conformations or active sites into the lowest free energy position, creating whole new predominant folding pathways or function.

It will not be easy to achieve a quantitative comparison of folding simulation and experiment beyond rates or other highly averaged quantities because many experimental probes are simply too difficult to compute from classical trajectories. Downhill folding is a good example: the complex fast interconversions among multiple states seen in simulations of WW domain 'underneath the surface' of folding kinetics[185, 186] are lumped into one 'molecular time' experimentally. One way to make progress is to find easily computed proxies for experimental reaction coordinates. For example, the tryptophan fluorescence wavelength of a large number of native and denatured proteins with variously exposed single tryptophans could be correlated with tryptophan side chain solvent exposure, local solvent electrostatics and other parameters easily extracted from simulation, to come up with reliable multi-parameter correlations between experiment and simulation. Pioneering attempts in this direction have been made already,[187] but were not fully successful in the past because of the enormous computational requirements,

now at hand. Conversely, it behooves experimentalists to develop faster ways to probe radius of gyration, FRET distance, and other variables that are easily and reliably computed. We are entering an era where simulations can run in a reasonable time frame the same protein with different labels attached or in different solvent conditions, so apples can be directly compared to apples. From the perspective of connecting simulation with experiment, protein science is entering an exciting time.

CHAPTER 2

REDUCING LAMBDA REPRESSOR TO THE CORE

2.1 Introduction*

Fragments of the lambda repressor protein containing amino acids 1-101 or 6-85 have long been subject to folding studies, including folding-function correlations,[188] remodeling of the hydrophobic core,[189] detection of fast two-state folding by NMR,[190, 191] and detection of multiple folding mechanisms of mutants.[192, 193] These fragments comprise as few as five of the helices from the complete protein, yet they have turned out to be thermodynamically quite stable in thermal and chemical denaturation studies.[194]

How small can a lambda repressor fragment be made while maintaining a high melting temperature and cooperativity? A previous experimental and computational analysis suggested that helices 1 and 4 constitute the minimal folding core and fold first.[195] Protein stability was particularly sensitive to mutations in this pair of helices, and molecular dynamics simulations showed that they are most likely to form secondary structure at room temperature. Explicit solvent replica-exchange thermodynamics simulations of a lambda repressor fragment (see Appendix A) also showed higher stability of helices 1 and 4 relative to other secondary structure. These results are in line with the idea that native-like preorganized secondary structure can accelerate folding of lambda repressor fragments,[196] whereas non-native secondary structure (e.g., extended structure that appears upon addition of guanidine hydrochloride[197]) can greatly slow down folding of lambda repressor fragment from microseconds to milliseconds.[198]

On the basis of this knowledge, we computationally redesigned λ^*_{6-85} (the Y22W mutant of fragment 6-85 of bacteriophage lambda repressor[199]) to

* This chapter is partially reproduced from Maxim B. Prigozhin, Krishnarjun Sarkar, Dennis Law, William C. Swope, Martin Gruebele and Jed Pitera, Journal of Physical Chemistry B, 115(9), 2090–2096, 2011, DOI: 10.1021/jp110175x

remove helices 2, 3, and 5, and connect helices 1 and 4 directly by a short flexible linker optimized to allow native-like contacts to form between these two helices. Two of the fragments, λ_{blue1} and λ_{blue2}, had a native-like modeled structure and proved most stable in molecular dynamics modeling, so we expressed them and investigated their folding experimentally. λ_{blue3} was also conformationally similar to λ^*_{6-85} but was energetically less favorable than the first two variants. It served as a control to reveal a correlation between the computed energies and the experimentally observed stability and folding rate.

λ_{blue1} turns out to have the same high melting point (62 °C) as the pseudo wild type λ^*_{6-85} fragment, and it folds at about the same rate ($\tau_{obs} = 15 \pm 4\ \mu$s). Its circular dichroism (CD) spectrum reveals well-formed helical content. Its melting points, as measured by CD and fluorescence, are in close agreement. λ_{blue1} is an apparent two-state folder. In contrast, its individual N and C terminal fragments are much less helical and do not show a cooperative transition. Thus, interaction between helices 1 and 4 is critical for the folding of λ_{blue1}. λ_{blue2} also has some helical content, while λ_{blue3} has a cooperative transition at much lower temperature, and a CD spectrum indicating more β-sheet content.

No attempt was made to redesign hydrophobic side chains that interacted with helices 2, 3, and 5 in the original λ^*_{6-85}, to keep the new miniprotein as close as possible to the original λ^*_{6-85} sequence. As a result, λ_{blue1} is more prone to aggregation than λ^*_{6-85}. We do predict two mutations to make the protein more soluble. Likewise, the fluorescence signal could be enhanced by adding a quencher (*e.g.*, histidine, cysteine) in contact with tryptophan 22. In its present form, λ_{blue1} will be useful for two applications: long time or replica exchange molecular dynamics simulations of fast folding that compare the reduced to the original molecule, and further redesign to create a stable two-helix bundle with low aggregation propensity and large fluorescence signal. We make PDB coordinates for the predicted structures available to facilitate such studies.

2.2 Experimental and Computational Methods

2.2.1 Simulation Protocol

All simulations were carried out using the AMBER10 suite of programs. Protein structures were visualized using the PyMOL software package. The AMBER parm96 all-atom force field[200] was used in conjunction with a modified generalized Born/solvent accessible surface area implicit solvent model.[201] The interior dielectric was 1.2, the exterior dielectric was 80.0, and the surface tension was 0.005 kcal/mol/Å^2. This combination of simulation parameters has been shown to be effective at discriminating between folded and unfolded conformations of many different proteins.[202]

After construction, the folded and unfolded forms of each variant protein were subjected to a two-stage structural optimization protocol. In the first stage, 10,000 steps of steepest descent energy minimization were performed while the atoms in helices 1 and 4 were restrained to their initial positions by relatively strong (5 kcal/mol/Å^2) Cartesian positional restraints. This allowed relaxation of the linker regions without distortion of the helices. In the second stage, an additional 10,000 steps of steepest descent energy minimization were performed without the helical positional restraints, allowing all atoms to move.

Once optimized, the folded and unfolded structures of each variant were simulated using conventional molecular dynamics at four different temperatures (300, 350, 400, 450 K). The folded structures of each variant were simulated for 6.25 ns, while the unfolded structures were simulated for 5 ns. The final 1.25 ns of both the folded and unfolded state trajectories were used for subsequent analysis. Uncertainties were estimated using the observed variance and correlation time of the data and are reported as ± 1 standard deviation. The target temperatures were enforced using an Andersen thermostat[203] with collisions every 10 ps. All bonds to hydrogen atoms were constrained to their equilibrium lengths using the SHAKE algorithm,[204] and a 2 fs time step was employed. The aggregate simulation time required for the entire project was 1.3 μs. Additional details are provided in the Supporting Information.

2.2.2 Protein Expression

λ_{blue1} gene inserted in a pET15-b vector with a histidine tag coding sequence was obtained from GenScript Corp. (Piscataway, NJ). λ_{blue2} and λ_{blue3} were derived from λ_{blue1} by site-directed mutagenesis (Quickchange kit; Stratagene, La Jolla, CA). Genes were expressed in Escherichia coli BL21 cells (Stratagene Corp., La Jolla, CA). The cell cultures were grown at 37 °C in LB Lennox Broth (Fisher Scientific Inc., Waltham, MA) and 100 μg/mL ampicillin (Fisher Scientific Inc., Waltham, MA) with shaking at 220 rpm. Protein biosynthesis was induced when absorbance at 600 nm reached 0.6–0.8 OD using isopropyl β-d-thiogalactopyranoside (IPTG) (Inalco, Milano, Italy) with the final concentration of 1 mM. Cell growth was allowed to proceed for 10 h after induction at 25 °C with agitation. The cells were collected by centrifugation for 10 min at 5000 rpm. Cell pellets were resuspended in 150 mL of the lysis buffer consisting of 50 mM Na_2HPO_4, 500 mM NaCl, and 10 mM imidazole at pH 8.0. Lysis was achieved by sonication on ice and cell debris was removed by centrifugation. The supernatant was loaded onto a Ni-NTA His-Bind column (Novagen Inc., Madison, WI) that was washed with lysis buffer. The column was then washed with a buffer containing 50 mM Na_2HPO_4, 500 mM NaCl, and 20 mM imidazole at pH 8.0 and eluted with another buffer containing 50 mM Na_2HPO_4, 500 mM NaCl, and 250 mM imidazole at pH 8.0. After purification, the proteins were dialyzed at 4 °C for two 6-hr periods against 0.1 M pH 7.0–7.5 phosphate buffer using 3500 MWCO dialysis tubing (Fisher Scientific Inc., Waltham, MA). The histidine tag was retained on all the mutants. Purity was confirmed by sodium dodecyl sulfate polyacrylamide gel electrophoresis, and the concentration was determined by UV-visible absorption spectroscopy.

2.2.3 Equilibrium Folding Experiments

Thermal unfolding experiments were performed by incrementing the temperature by 3 °C in the 20–89 °C range and measuring CD at 222 nm as well as the integrated fluorescence with the 280 nm excitation wavelength at each temperature using a Jasco spectrometer equipped with a Peltier temperature control (Jasco Inc., Easton, MD). Output fluorescence was passed through a long-pass filter with FWHM transmittance at 305 nm. CD spectra

in the 200–250 nm region were recorded at 20 °C using the same spectrometer. The concentration of protein solutions used in CD measurements was approximately 10 μM.

2.2.4 Temperature Jump Experiment

Temperature jumps on the order of 8–10 °C were initiated with a Surelite Q-switched Nd:YAG laser (Continuum Inc., Santa Clara, CA) Raman-shifted to 1.9 μm. Fluorescence excitation was achieved by a Mira Ti:Sapphire laser (Coherent Inc., Santa Clara, CA) tripled to the wavelength of 281 $\pm$ 2 nm using a third harmonic generator (CSK Optronics Inc., Torrance, CA). The power of the laser at the sample cell was usually on the order of 3 mW at 14 ns time intervals between pulses. Peptide fluorescence was passed through a B370 band-pass filter from Hoya Corp. (Santa Clara, CA). The data were digitized using a 500 ps resolution oscilloscope (Tektronix RTD720A; Tektronix Inc., Beaverton, OR) and analyzed with a program written in LabWindows (National Instruments Inc., Austin, TX). Equilibrium temperature of the sample was maintained using an automated temperature controller (Lake Shore 330; Lake Shore Cryotronics Inc., Westerville, OH). The details of the T-jump instrument can be found elsewhere.[205] The jump temperature was calibrated using a 250 μM solution of *L*-tryptophan (Sigma-Aldrich Inc., St. Louis, MO). Sample concentrations were difficult to measure due to protein aggregation. On the basis of the T-jump signal and absorbance measurements, the samples were in the concentration range of 20–30 μM.

2.2.5 Data Analysis

All data were analyzed using an Igor Pro software package (Wavemetrics Inc., Lake Oswego, OR). Fluorescence decays were fitted to a linear combination of the lifetime right before the jump ($\chi = 1$) and a lifetime 0.5 ms after the jump ($\chi = 0$). This analysis allowed to plot relative lifetime shift as a function of time, $\chi(t)$. The zero time was estimated from the Raman scattering peak of the solvent and the first five decays were not included in the analysis (Figure 2.5). The resulting exponential decay was fitted to obtain the folding-unfolding relaxation time constant τ. Thermal denaturation curves

from both fluorescence and CD experiments were fitted using a two-state thermodynamic model. Fluorescence signal base lines $F_i(T)$ were assumed to be linear functions of temperature for the native and denatured states $i = N$ and D:

$$F_i = b_i + m_i(T - T_m) \tag{2.1}$$

b_i and m_i were obtained by fitting the first and the last 3–4 points in the CD and fluorescence melt profiles. The overall signal $F(T)$ was then fitted as a linear combination of the fraction that each population contributed at temperature T multiplied by the baseline signal of that population (equation 2.2).

$$F(T) = \frac{F_D}{1 + K_{eq}} + \frac{F_N K_{eq}}{1 + K_{eq}}, \text{ with } K_{eq} = e^{-\Delta G(T)/RT} \tag{2.2}$$

The free enegy was expanded as a Taylor series about the transition temperature between folded and unfolded states (equation 2.3):

$$\Delta G(T) \approx \Delta G(T_m) + \Delta S(T_m)T_m - \Delta S(T_m)T \tag{2.3}$$

The $\Delta G(T_m)$ term is equal to zero because the difference in free energy between folded and unfolded states at the midpoint temperature T_m is zero. The linear coefficient $\Delta S(T_m) = -\delta\Delta G(T)/\delta T$ would correspond to the folding entropy at T_m in the limit where equation 2.3 is exact over the entire fitting range. A second order term in the Taylor expansion was not needed to fit the data within measurement uncertainty. An alternative model for ΔG in terms of a constant heating capacity for folding gave a similar quality fit and T_m.

2.3 Results

2.3.1 Design of Truncated Protein Structures

The initial structure used for the construction of truncated lambda repressor variants was taken from the 1LMB crystal structure,[206, 207] with substitution Y22W in helix 1 corresponding to a fluorescent pseudo wild type

λ^*_{6-85}.[196, 198, 199] An additional mutation D14A slightly stabilizes (2.5 °C) the full λ^*_{6-85} fragment.[196] The truncated proteins were constructed using portions of the two largest helices, 1 and 4, connected by short glycine-serine linker segments.

Three variants (a–c) of each helix were constructed (Table 2.1). The sequences were chosen to minimize the size of helix 1 while preserving its interactions with helix 4. In h1a and h1b, the residues linking helix 1 to helix 2 (Gly 30 to Ser 32) were retained as part of the linker to help connect helix 1 to helix 4 in the modified protein. In h1b an additional tyrosine was inserted, which interacts strongly with Trp 22 in some λ^*_{6-85} mutants.[198] Three variants of helix 4 were constructed in a similar fashion (Table 2.1). The residues that connect helices 3 and 4 in λ^*_{6-85} (Phe 51/Gly 53 to Leu 57) were kept to form part of the linker between helix 1 and helix 4.

Table 2.1: Residue ranges and sequences of lambda repressor helix variants

Variant name	λ^*_{6-85} residue range	Amino acid sequence
h1a	7-32	LTQEQLEAARRLKAIWEKKKNELGLS
h1b	7-33	LTQEQLEAARRLKAIWEKKKNELGLSY
h1c	7-30	LTQEQLEAARRLKAIWEKKKNELG
h4a	51-71	FNGINALNAYNAALLAKILKV
h4b	51-77	FNGINALNAYNAALLAKILKVSVEEFS
h4c	53-70	GINALNAYNAALLAKILK

Visual inspection of the truncated helix pairs suggested that the C-terminus of the helix 1 variants would be separated from the N-terminus of the helix 4 variants by 10–20 Å, so a series of three extended Gly-Ser repeat linkers (GSG, GSGS, GSGSG) were tested. The 27 possible helix 1/linker/helix 4 combinations were then constructed in a putative folded and in a fully extended conformation (see Computational and Experimental Methods). As a reference, two variants with the wild type sequence linking helices 1 and 4 instead of the Ser-Gly linkers were also constructed.

2.3.2 Simulation to find low energy structures

All 27 variants and two controls were subjected to the same automated protocol of structural optimization by energy minimization using implicit solvent simulations to provide a coarse measure of the stability of each sequence in the

folded state (see Computational and Experimental Methods and Appendix A for details). During the first stage of optimization, the atoms in helices 1 and 4 were restrained to their initial positions. This allowed relaxation of the linker regions without distortion of the helices. In the second stage, all atoms were allowed to move. Once structurally optimized, the folded and unfolded conformations of each variant were simulated at 300 K using conventional molecular dynamics. Additional simulations were carried out at 350, 400, and 450 K as controls but were not used to select sequences for synthesis, since most fragments unfolded at the higher temperatures. Ideally, large-scale thermodynamic sampling such as replica-exchange molecular dynamics would be performed for each variant to estimate its stable structures and melting temperature. However, our experience is that such simulations require extensive simulation to converge for proteins as large as lambda repressor, making them impractical for the screening approach taken here.(8) Though limited, our simulations were sufficient to allow local structural relaxation of the linker and interhelical angles in folded conformations and to allow the extended 'unfolded' conformations to collapse to disordered structures with radii of gyration comparable to the folded conformations. The average radius of gyration of the folded simulations is 11.1 ± 0.1 Å compared to 11.3 ± 0.7 Å for the unfolded simulations. Two metrics were used to assess the stability of the folded state for each variant sequence (Table 2.2). First, the C_α-RMSD of the residues in helices 1 and 4 (relative to the original X-ray crystal structure) was calculated for each saved coordinate. Sequences were classified as structurally stable if the average C_α-RMSD for the helices was less than 2.5 Å. Second, the per-atom energy gap between folded and unfolded conformations at 300 K was used as a measure of energetic stability, allowing comparison of sequences with different lengths. The mutants with the wild type linker (bottom of Table 2.2) had energy gaps of –0.023 to –0.025 kcal/mol/atom. Thus sequences were classified as energetically stable if the observed energy gap was at least –0.020 kcal/mol/atom.

2.3.3 Protein Selection and Expression

Three variants from Table 2.2 with low C_α-RMSD values suggesting structural stability were chosen for expression. The two sequences λ_{blue1} = h1a(GSGSG)h4b

Table 2.2: Twenty nine truncated variants, including two wild type linkers for reference. Average folded and unfolded energies, energy difference per atom, and average C_α-RMSD for helices 1 and 4 at 300 K are shown. Reported uncertainty estimates are one standard deviation. Italic entries were expressed and studied experimentally.

Variant name	$\langle E_{folded} \rangle$, kcal/mol	$\langle E_{unfolded} \rangle$, kcal/mol	$\Delta E/atom$, kcal/mol	$\langle$RMSD$\rangle$ (Å)
h1aGSGh4a	-1211.3 ± 1.7	-1159.3 ± 2.7	-0.065 ± 0.004	3.35 ± 0.14
h1aGSGh4b	-1322.6 ± 2.0	-1309.8 ± 3.2	-0.014 ± 0.004	4.93 ± 0.05
h1aGSGh4c	-1135.0 ± 1.9	-1134.7 ± 3.6	-0.000 ± 0.005	3.08 ± 0.07
h1aGSGSh4a	-1209.4 ± 3.2	-1186.7 ± 2.8	-0.028 ± 0.005	4.08 ± 0.08
h1aGSGSh4b	-1324.4 ± 2.2	-1303.9 ± 2.1	-0.023 ± 0.003	1.80 ± 0.07
h1aGSGSh4c	-1119.1 ± 2.5	-1097.6 ± 2.6	-0.028 ± 0.005	2.87 ± 0.14
h1aGSGSGh4a	-1198.1 ± 3.4	-1158.6 ± 3.7	-0.048 ± 0.006	3.17 ± 0.17
h1aGSGSGh4b	-1325.5 ± 4.4	-1301.5 ± 3.3	-0.026 ± 0.006	2.38 ± 0.14
h1aGSGSGh4c	-1125.7 ± 1.6	-1093.8 ± 4.0	-0.041 ± 0.006	3.40 ± 0.15
h1bGSGh4a	-1210.4 ± 2.4	-1186.8 ± 2.8	-0.029 ± 0.004	2.94 ± 0.07
h1bGSGh4b	-1318.2 ± 2.0	-1311.7 ± 2.8	-0.007 ± 0.004	3.43 ± 0.23
h1bGSGh4c	-1135.0 ± 1.7	-1124.9 ± 2.9	-0.013 ± 0.004	2.83 ± 0.19
h1bGSGSh4a	-1212.8 ± 4.2	-1195.1 ± 5.3	-0.021 ± 0.008	5.07 ± 0.10
h1bGSGSh4b	-1343.2 ± 3.1	-1331.0 ± 2.0	-0.013 ± 0.004	1.95 ± 0.12
h1bGSGSh4c	-1142.7 ± 1.7	-1110.6 ± 3.5	-0.041 ± 0.005	4.73 ± 0.07
h1bGSGSGh4a	-1211.2 ± 1.7	-1186.0 ± 2.7	-0.030 ± 0.004	6.11 ± 0.16
h1bGSGSGh4b	-1328.3 ± 3.3	-1325.8 ± 3.3	-0.003 ± 0.005	3.07 ± 0.03
h1bGSGSGh4c	-1137.5 ± 1.9	-1120.8 ± 2.0	-0.021 ± 0.003	3.28 ± 0.15
h1cGSGh4a	-1187.9 ± 2.8	-1174.4 ± 4.6	-0.017 ± 0.007	2.47 ± 0.05
h1cGSGh4b	-1311.2 ± 4.2	-1284.7 ± 4.9	-0.031 ± 0.006	2.06 ± 0.07
h1cGSGh4c	-1115.4 ± 3.1	-1096.9 ± 2.9	-0.026 ± 0.006	3.18 ± 0.07
h1cGSGSh4a	-1207.4 ± 2.6	-1160.7 ± 2.0	-0.060 ± 0.004	3.17 ± 0.09
h1cGSGSh4b	-1298.2 ± 2.8	-1296.2 ± 3.8	-0.002 ± 0.005	3.11 ± 0.13
h1cGSGSh4c	-1122.5 ± 2.7	-1114.6 ± 3.0	-0.011 ± 0.006	3.30 ± 0.06
h1cGSGSGh4a	-1174.0 ± 3.0	-1168.8 ± 1.9	-0.007 ± 0.005	4.17 ± 0.05
h1cGSGSGh4b	-1297.7 ± 2.4	-1274.3 ± 3.6	-0.027 ± 0.005	3.69 ± 0.46
h1cGSGSGh4c	-1114.0 ± 2.5	-1092.9 ± 2.3	-0.029 ± 0.004	5.21 ± 0.05
h1aWTh4a	-1401.4 ± 3.7	-1377.9 ± 3.9	-0.023 ± 0.005	3.15 ± 0.12
h1aWTh4b	-1536.6 ± 3.5	-1509.3 ± 2.6	-0.025 ± 0.004	1.98 ± 0.02

and λ_{blue2} = h1a(GSGS)h4b had the lowest folded energy and also satisfied the stability criterion, having per-atom energy gaps of at least -0.020 kcal/mol/atom. The third sequence, λ_{blue3} = h1b(GSGS)h4b, had only half the energy gap per atom. We synthesized it as a control to see if computed stability gaps and experimental melting points would correlate. Figure 2.1 shows the computed structures of the three chosen variants.

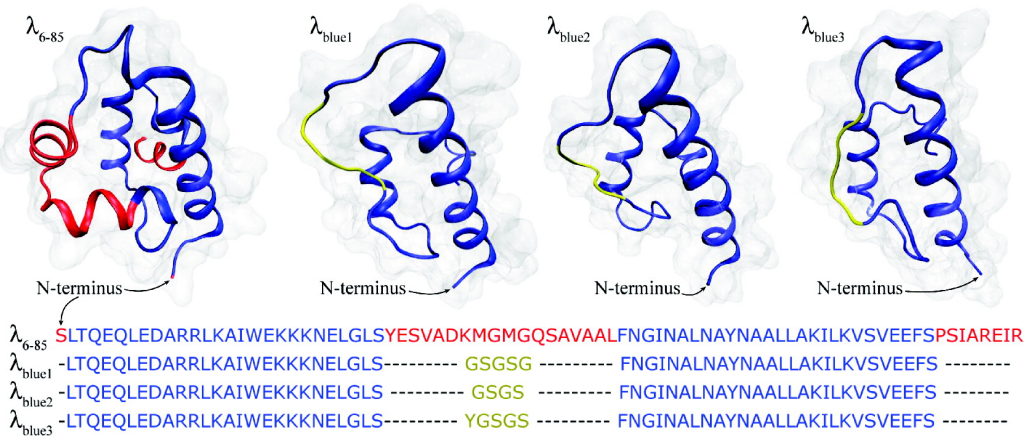


Figure 2.1: Model structure of λ^*_{6-85} based on the X-ray crystal structure of a closely related lambda repressor fragment PDB 3KZ3,(5, 20) and molecular dynamics models of the structures for $\lambda_{blue1-3}$ computed in the present work. A van der Waals surface is also shown for each model, along with their sequences.

For protein expression, the λ_{blue1} gene was inserted in a pET15-b vector, and the protein was expressed as described in Computational and Experimental Methods. The other two variants were made by site-directed mutagenesis and expressed similarly. A fourth variant (h1cGSGh4b) was also attempted but did not express well, possibly because the higher hydrophobicity of the longer fragment induced more severe proteolysis in the expression host. Expression of the designed protein thus becomes an additional practical selection criterion.

2.3.4 λ_{blue1} Has Helical Structure and Unfolds Cooperatively

The three variants we expressed were sufficiently soluble to carry out equilibrium experiments. CD spectra for each λ variant were collected at 20 °C from 200 to 250 nm (Figure 2). The CD spectrum of λ_{blue1} had the most negative ellipticity. The CD profile of λ_{blue1} showed characteristic α-helix peaks

at 209 and 222 nm. The deeper peak at 222 nm is indicative of a high helix to coil ratio. In contrast, the CD profile of λ_{blue2} had a deeper peak at 210 nm, indicative of increased random coil. The small CD signal of λ_{blue3} with a single minimum at 218 nm had no α-helical signature. The mean residue ellipticity of λ_{blue1} was at least –8000° M^{-1} m^{-1} at 222 nm (Computational and Experimental Methods).

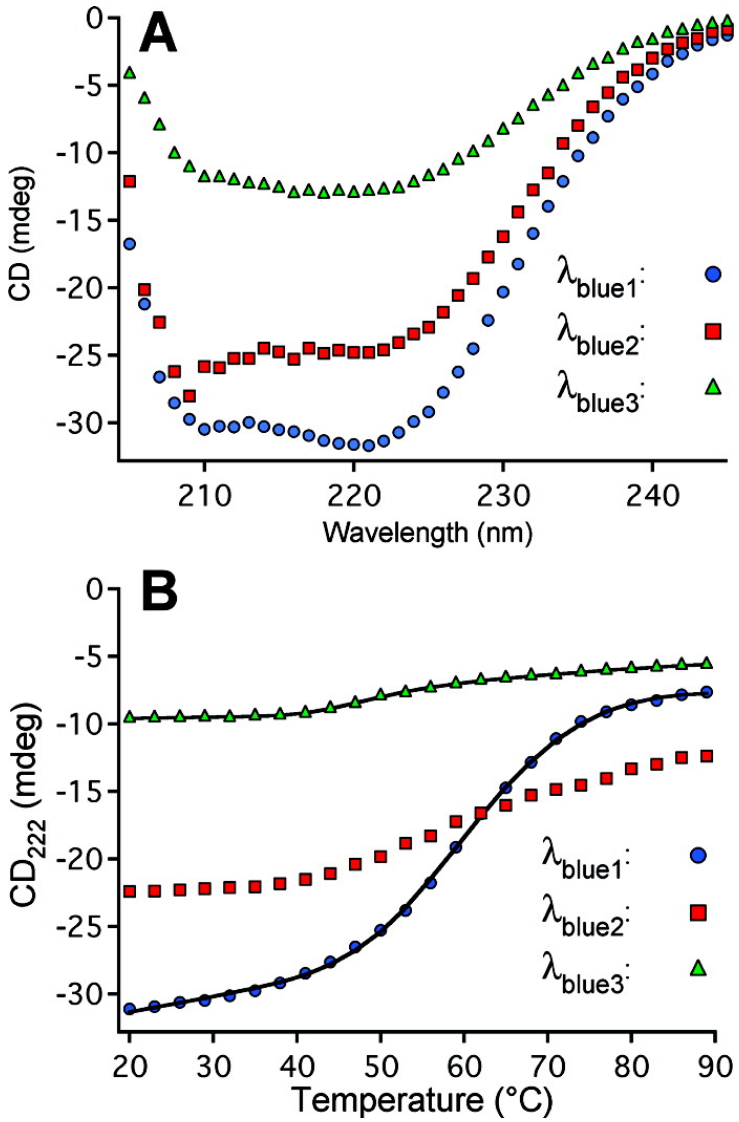


Figure 2.2: Circular dichroism spectra (A) and thermal denaturations for $\lambda_{blue1-3}$. Thermodynamic fits are shown as black lines in the bottom plot, with values for the fitting parameters in Table 2.3.

Thermal unfolding of the three constructs was measured by circular dichroism at 222 nm, as shown in Figure 2.2 (Computational and Experimental

Methods). The CD melt of λ_{blue1} exhibited a cooperative melting transition with a melting temperature of $T_m = 62 \pm 2$ °C and $\Delta S(T_m) = 0.41 \pm 0.01$ kJ/mol/K (Table 2.3 and Computational and Experimental Methods). The CD melt of λ_{blue3} yielded a T_m of 47 ± 3 °C but similar cooperativity. The CD melt for λ_{blue2} could not be fitted accurately because of the high temperature baseline, but visual inspection of Figure 2.2 indicates that it lies between the other two molecules.

Table 2.3: Fitting results for the thermodynamic titration experiments (± errors are two standard deviations of the mean)

		Protein	
Experiment	Parameters	λ_{blue1}	λ_{blue3}
CD melt	T_m (°C)	62 ± 2	47 ± 3
	$\Delta S(T_m)$ (kJ/mol °C)	0.406 ± 0.012	0.57 ± 0.16
Fluorescence	T_m (°C)	65 ± 2	45 ± 2
melt	$\Delta S(T_m)$ (kJ/mol °C)	0.406 (fixed)	0.47 ± 0.08

Thermal unfolding was also measured by integrated fluorescence intensity, as shown in Figure 2.3 (Computational and Experimental Methods). Lambda repressor fragments without the Q33Y residue are known to have only a small change in tryptophan fluorescence upon folding. For λ_{blue1} the transition could be observed (Figure 2.2 and inset), with $T_m = 65 \pm 2$ °C. This melting temperature differs by 3 °C from the CD measurement, but considering the very small fluorescence signal and uncertainties, this difference is probably not significant. For λ_{blue2}, the transition lies at about 50 °C. The λ_{blue3} sequence contains the Q33Y mutation, and a much larger fluorescence change is observed, indicating that despite the nonhelical CD spectrum, an interaction between W22 and Y33 is formed upon folding. The melting temperature is 45 ± 2 °C, within measurement uncertainty of the CD transition midpoint.

2.3.5 λ_{blue1} fragments are less helical and not cooperative

The above results beg the question whether the apparent cooperativity and secondary structure is really a property of the whole miniprotein, or just independent structure and folding of helices 1 and 4. We purchased fragments 1–26 (helix 1) and 32–58 (helix 4) of λ_{blue1} and measured their CD spectra and

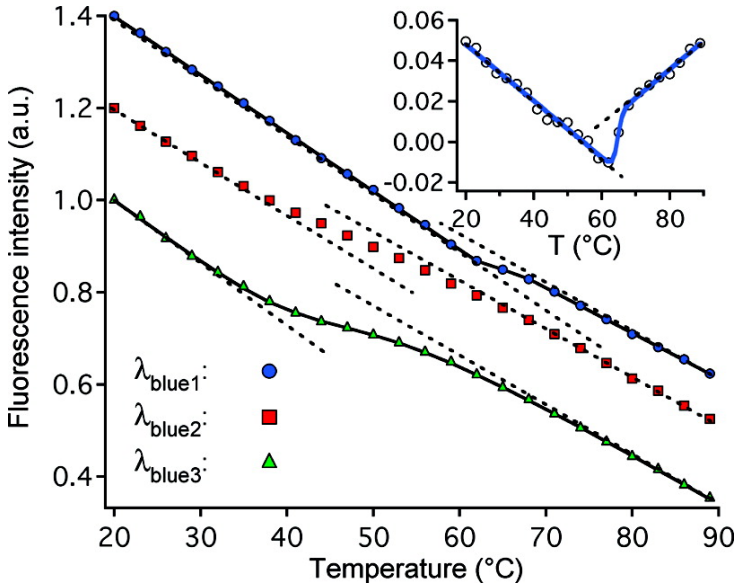


Figure 2.3: Fluorescence intensity-detected thermal denaturation traces for $\lambda_{blue1-3}$. Thermodynamic fits are shown as black lines in the bottom plot, with values for the fitting parameters in Table 2.3. The inset shows the λ_{blue1} transition with the average overall baseline subtracted to reveal the transition more clearly.

thermal titration. Both fragments show much deeper peaks at 205–208 nm in the CD spectrum, indicative of more random coil content (Figure 2.4). Both fragments, as well as their 1:1 mixture, show no cooperative CD transition. Fragment 1–26 has a broad transition at much lower temperature than the full miniprotein. Most notably, a mixture of the two fragments shows no enhanced cooperativity, indicating that the entropic constraint provided by the linker is important for folding.

2.3.6 Temperature jump experiments

Since λ_{blue1} had a CD spectrum consistent with helical structure, a melting point similar to the pseudo wild type, and a cooperative fluorescence transition, we performed a laser temperature jump experiment. Protein solutions (20–30 μM) were jumped by 8–10 °C, and relaxation kinetics were detected by the picosecond time-resolved change in fluorescence decay (see Computational and Experimental Methods). We observed a relaxation phase that could be fitted to $\tau_{obs} = 15 \pm 4$ μs at a final temperature of 63 ± 1 °C. No further relaxation was observed up to 0.5 ms (Figure 2.5). Kinetics of λ_{blue2}

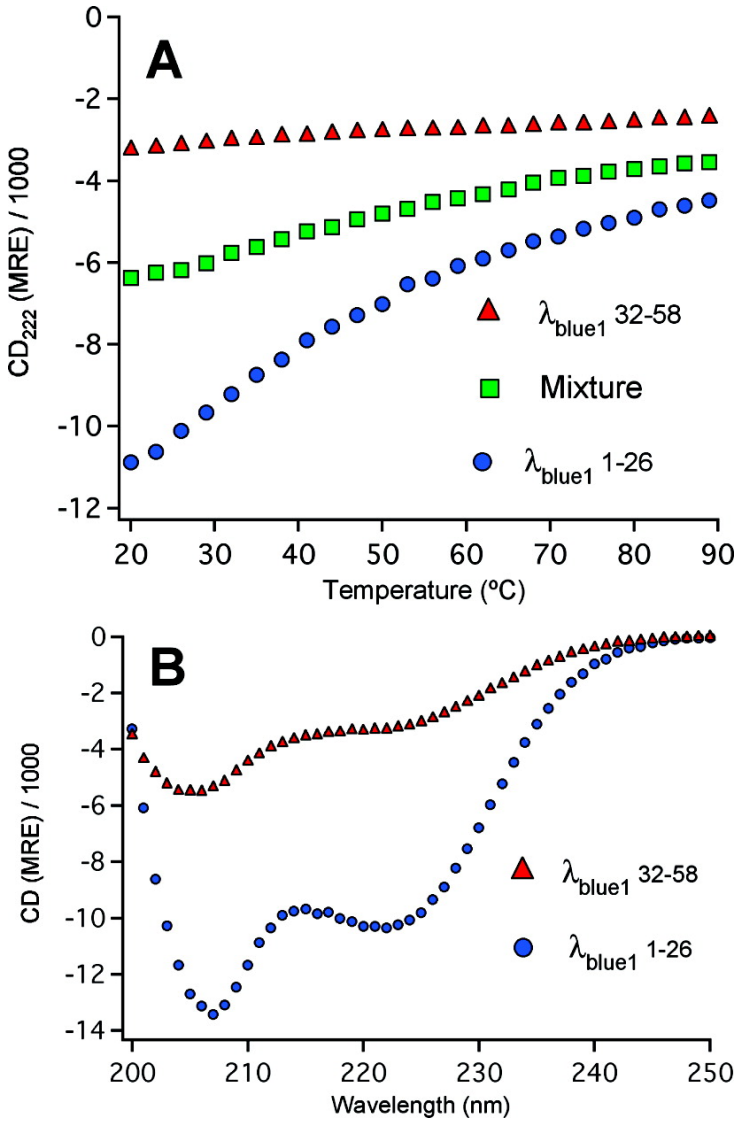


Figure 2.4: CD spectra and thermal melts of N terminal and C terminal fragments of λ_{blue1}. The mixture of both peptides in A simply produces an average of the two individual fragment melting curves. The spectra in B show much more random coil content (negative CD below 210 nm) than full λ_{blue1} in Figure 2.2.

folding could not be resolved. λ_{blue3} showed a single fast phase (1.3 ± 0.4 μs) when jumped to a final temperature of 49 ± 1 °C.

2.4 Discussion and conclusion

As expected based on the computed per atom energy gap between the folded and unfolded states, the control λ_{blue3} turned out to be considerably less stable than λ_{blue1} and λ_{blue2}. Although it had one of the smallest computed C_α-RMSD values (Table 2.2), it has low helix content and the lowest melting point (45–47 °C) observed among the three constructs. The small difference in the computed energy gaps of λ_{blue1} and λ_{blue2} did not enable them to be ranked correctly, but λ_{blue2} had the larger computed RMSD and turned out to be less helical than λ_{blue1} by CD spectroscopy.

Of the three peptides studied, only λ_{blue1} had all of the features of a small helix bundle protein: cooperative CD and fluorescence melts, a CD spectrum with a minimum at 222 nm deeper than at 210 nm, a high T_m, and a large cooperativity parameter ΔG_1. The full λ^*_{6-85} has a melting temperature of 61 °C under the same solvent conditions, and a fitted $\Delta S(T_m)$ = 0.68 kJ/mol/K.[198] In Table 2.3, we find 62 °C and 0.41 kJ/mol/K for λ_{blue1}. The slightly smaller cooperativity is expected for a smaller molecule. Based on a simple model of linear scaling of protein heat capacity with chain length,[208] one would predict $\Delta S(\lambda_{blue1}) = \Delta S(\lambda^*_{6-85})\mathrm{N}(\lambda_{blue1})/\mathrm{N}(\lambda^*_{6-85}) \approx$ 0.49 kJ/mol/K. Thus, λ_{blue1} is nearly as stable and cooperative a folder as λ^*_{6-85} when correcting for the smaller protein size. Moreover, the results for isolated helices shown in Figure 4 prove that helical structure and cooperative folding are really properties of the miniprotein, not of the N- and C-terminal helices alone.

The folding relaxation time of λ_{blue1} near its melting temperature (15 μs at 63 °C) is also very close to λ^*_{6-85} (17 μs at 61 °C). The linker connecting helices 1 and 4 in λ_{blue1} is shorter than in the pseudo wild type (5 *vs.* 18 residues). On the basis of entropic considerations alone (random chain linkers), one would have expected λ_{blue1} to fold slightly faster than λ^*_{6-85}. The comparable relaxation times indicate one of two things: either the contact energy driving term of λ_{blue1} is smaller than that of the pseudo wild type, or helices 2 and 3 in the wild type form a structured and compact linker during

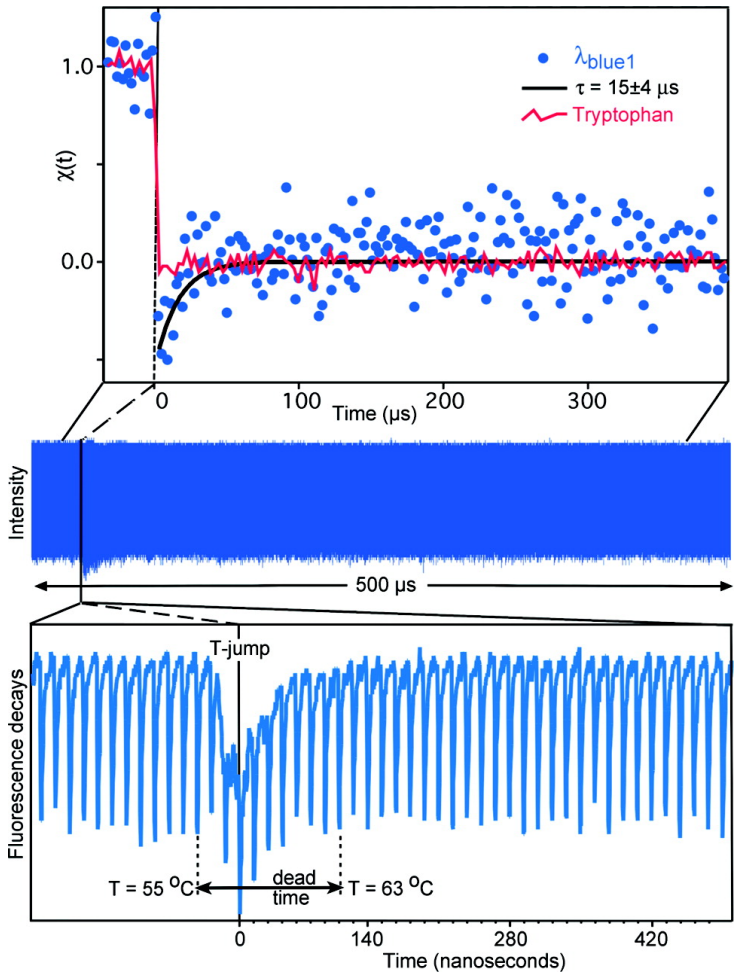


Figure 2.5: Temperature jump experiment: 0.5 ms of data containing over 30,000 tryptophan fluorescence decays were collected (the blue band shows fluorescence intensity), with a detail of the raw data near the T-jump shown below. Raman scatter from the infrared pump beam obscures a few fluorescence decays, setting the dead time of ≈50 ns. The top graph shows the fluorescence lifetime analysis $\chi(t)$ for λ_{blue1} (15 μs fitted response) and for tryptophan (instantaneous response). The 15 μs relaxation phase is also visible in the intensity data near $t = 0$.

helix 1–4 folding.

The nearly identical stability and fast folding of λ_{blue1} relative to λ^*_{6-85} suggest that the formation and docking of helices 1 and 4 is both the principal thermodynamic stabilizing interaction and kinetic docking step during folding. This obervation confirms the mutation and molecular dynamics results that led to the present work,[195] which suggested that helices 1 and 4 are the earliest to form and make contacts. Oas and co-workers suggested a transition state without helix 3 in the rate-limiting step,[209] and our result is consistent with that analysis. However, it appears that helix 2 is not as important in the early kinetics and contact formation as suggested by their collision-diffusion model.

Helices 1 and 4 lie close to the N and C termini of the full λ^*_{6-85} fragment. A number of small proteins form initial contacts between two remote helices separated by long loops (*e.g.*, apomyoglobin[210] and cytochrome *c*[211]). The rapid folding kinetics of λ^*_{6-85}, and identification of helices 1 and 4 as its folding core, together suggest that early contact formation between two remote helices is also predominant there. As long as the greater reduction of loop entropy during collision of remote helices is compensated by optimal contact energies between these helices, the free energy penalty associated with this mode of structure formation can be ameliorated. It remains unclear why the 'remote' mechanism should be preferred in vitro over a mechanism where adjacent helices initiate folding. We speculate that there could be a reason in vivo. A small contact energy between 'center helices' and a large contact energy between 'outer' helices could prevent early collapse and non-native contact formation until an entire protein subunit is available for folding. There is certainly evidence in some cases for sequence-control during cotranslational folding to improve in vivo foldability of proteins.[212] Such an explanation is appealing for whole proteins (*e.g.*, apomyoglobin, apocytchrome *c*) but could also apply to internal subunits of proteins (*e.g.*, lambda repressor) if they fold independenly, which λ^*_{6-85} and λ_{1-101} clearly do.

With a relaxation time of $\tau \geq 15$ μs, the reduced lambda repressor is still an approximate two-state folder by the criterion of fast/slow kinetic phase amplitudes from ref [213]. However, it is not far from being an incipient downhill folder with a barrier of <3 k_BT based on a survey of many lambda repressor mutants.[192] In the present work, we left hydrophobic residues in

helices 1 and 4 unmodified even if they were no longer required for native contact formation. It may be possible to increase the folding rate further by reducing non-native hydrophobic contacts. It certainly should be possible to improve solubility by eliminating some of the most solvent-exposed hydrophobic side chains. We predict Y60 and L64 on helix 4 as prime candidates for replacement by similarly sized polar residues. However, for comparison with molecular dynamics simulations contrasting a reduced folding core with a full small protein, λ_{blue1} is the best target as it most closely resembles the sequence of the original protein.

CHAPTER 3

THE FAST AND THE SLOW: FOLDING AND TRAPPING OF λ_{6-85}

3.1 Introduction*

Small proteins have become important benchmarks for folding theory[214] and simulation.[33] Single- and multitrajectory simulations now make easily testable predictions about folding rates and stabilities.[33, 215, 216] Lambda repressor fragments, originally a paradigm for stability measurements[217] and later adapted for NMR and temperature-jump (T-jump) kinetics,[30, 218] have become the largest "small" platforms for computational studies:[33, 219] 80 amino acids in five helices for the λ_{6-85} fragment. Experimentally, slower-folding mutants are assigned as apparent two-state folders,[31, 218] whereas sub-20 μs mutants are assigned as downhill or incipient downhill folders.[30, 220] The latter assignment is based on the 2 μs "molecular phase" observed only upon stabilization of the native state and attributed to direct observation of the barrier crossing.[30, 221] A recent long-trajectory simulation confirmed the low barriers determined experimentally (1.5 k_BT for λD14A in Table 3.1).[219]

Experimental analyses of two reaction coordinates[31] and multiple probes[222] using alanine-rich and -poor mutants of λ_{6-85} demonstrated that even apparent two-state folders require $\geq$2 reaction coordinates and >2 states for a full description. In addition, λ_{6-85} has a propensity to adopt an extended structure (here abbreviated PP/β because it is found in polyproline helices or β strands) at high temperature and in mild denaturant.[223]

Indeed, a rich energy landscape structure has been predicted by a Markov state analysis of multiple λ_{6-85} folding trajectories. For the λD14A mutant (Table 3.1), Bowman *et al.*[33] carried out relaxation simulations start-

* This chapter is partially reproduced from Maxim B. Prigozhin and Martin Gruebele, Journal of the American Chemical Society, 133(48), 19338–19341, 2011, DOI: 10.1021/ja209073z

ing with equal populations in the denatured microstates. They found two <10 μs phases, in general agreement with T-jump experiments, but also a much slower ≥1 ms phase. The slow phase is present in simulations without denaturant and is associated with compact β-rich denatured states; it disappears only when the β structural ensemble is removed from the calculation.

Bowman *et al.* offered several possible explanations for the calculated slow phase: it may have been missed by laser T-jump experiments, which usually extend to <0.5 ms; the slow phase may represent very slow folding of λ_{6-85} in denaturant, as previously seen by experiment;[224] or the force field they used may overemphasize β-sheet stability.

We have continued the experimentsimulation dialogue by carrying out new 5 ms T-jump measurements on alanine-rich and -poor mutants of λ_{6-85}. We show that all of the above explanations apply: We detected no significant slow phase for λD14A, so the force field or the equal-population initial condition overestimates the β-sheet propensity for this mutant. However, we detected a slow phase for two even more stable λ_{6-85} mutants, λQ33Y and λsQ33Y, proving that slow interconversion between structural ensembles of λ_{6-85} can occur without denaturant. Moreover, λD14A also begins to show a slow phase in mild denaturant. The computational suggestion that compact β-rich traps result in slow dynamics is thus plausible, although λ_{6-85} still refolds in microseconds from the extensively unfolded state if it can avoid such traps. Our results highlight once again the importance of thermodynamic tuning when comparing experiment and computation, because small free energy discrepancies (δG) can produce large population differences ($e^{-\delta G/RT}$).

3.2 Materials and methods

3.2.1 Sample preparation

λD14A was expressed as described previously.[225] λQ33Y, λnQ33Y, λnQ33Y, and λsA49G were also expressed according to the published procedures.[31] Samples were lyophilized and the measurements were done in 50 mM K_3PO_4 buffer at pH = 7.1. Guanidine hydrochloride (GuHCl) and *L*-tryptophan were obtained from Sigma-Aldrich Inc. (St. Louis, MO).

3.2.2 Equilibrium folding experiments

Circular dichroism (CD) thermal melts (Figures B.1–B.6, Table B.1) were done on a Jasco-715 spectropolarimeter (Jasco Inc., Easton, MD). The temperature was incremented by 3 °C from the initial value of 20 °C until the final temperature of 95 °C was reached resulting in 26 CD spectra. Temperature was controlled using a Peltier element. Each CD spectrum in figures is an average of 5 spectra at 100 nm/min scan speed. A quartz cuvette (Starna Cells Inc., Atascadero, CA) with 1 cm path length was used. All proteins were used at 2.5 μM concentration. Fluorescence spectra (Figures B.1–B.6, Table B.1) were measured using a fluorescence spectrophotometer Cary Eclipse (Varian Inc., Palo Alto, CA). The PMT voltage was usually 640-660 V, excitation and emission slit widths were 5 nm each, excitation wavelength was 280 nm, interval between points was 1 nm, and the scan rate was 600 nm/min. Sample concentrations were 10 μM for all proteins.

3.2.3 T-jump setup

Laser temperature jump was done using a Surelite Q-switched Nd:YAG laser (Continuum Inc., Santa Clara, CA) Raman-shifted to 1.9 μm by passing the beam through a 1 m long tube with hydrogen gas pressurized to 300 psi. (Figures 3.1 and B.7) The beam was then passed through a 50% beam splitter so that the sample could be excited from two sides, which provided for more uniform heating of the sample during the temperature jump. The jump was usually 5-7 C. The exact size of the jump was calibrated by comparing the fluorescence decays of tryptophan (300 μM solution) after the jump with the corresponding decays without the jump at the equilibrium temperature several degrees higher. Equilibrium temperature was set using an automated temperature controller (Lake Shore 330, Lake Shore Cryotronics Inc., Westerville, OH). The sample cell was made of fused silica tubing (3530S-100, VitroCom, Mountain Lakes, NJ) welded on one side. The excitation path length was 0.3 mm. The sample was excited with a Ti:sapphire laser (KMLabs Inc., Boulder, CO), which pulsed at 80 MHz frequency. The laser wavelength was 860 nm, which was frequency tripled with a third harmonic generator (CSK Optronics Inc., Torrance, CA) to 287 nm. Tryptophan fluorescence was then guided with an optical cable, passed through a B370 band-pass filter (Hoya

Corp., Santa Clara, CA) and collected by a photomultiplier (R7400U-03, Hamamatsu Corp., Bridgewater, NJ). The signal was then recorded and digitized with an oscilloscope (DPO7254, Tektronix Inc., Beaverton, OR) with 2.5 GHz bandwidth. The length of the time traces was 5 ms and each trace contained 50 000 000 data points. Temperature jump was set to occur 153.75 μs after the oscilloscope was triggered to start data collection. The sampling frequency was 10 GS/s, which yielded a data point every 100 ps. The signal was usually 100-250 mV. The data were analyzed with software written in LabView (National Instruments Inc., Austin, TX) and MATLAB (Mathworks Inc., Natick, MA). Sample concentrations were 300 μM, 225 μM, 364 μM, 281 μM, 385 μM, 323 μM for Trp, λQ33Y, λsQ33Y, λD14A, λnQ33Y, and λsA49G, respectively, as measured by absorption at 280 nm.

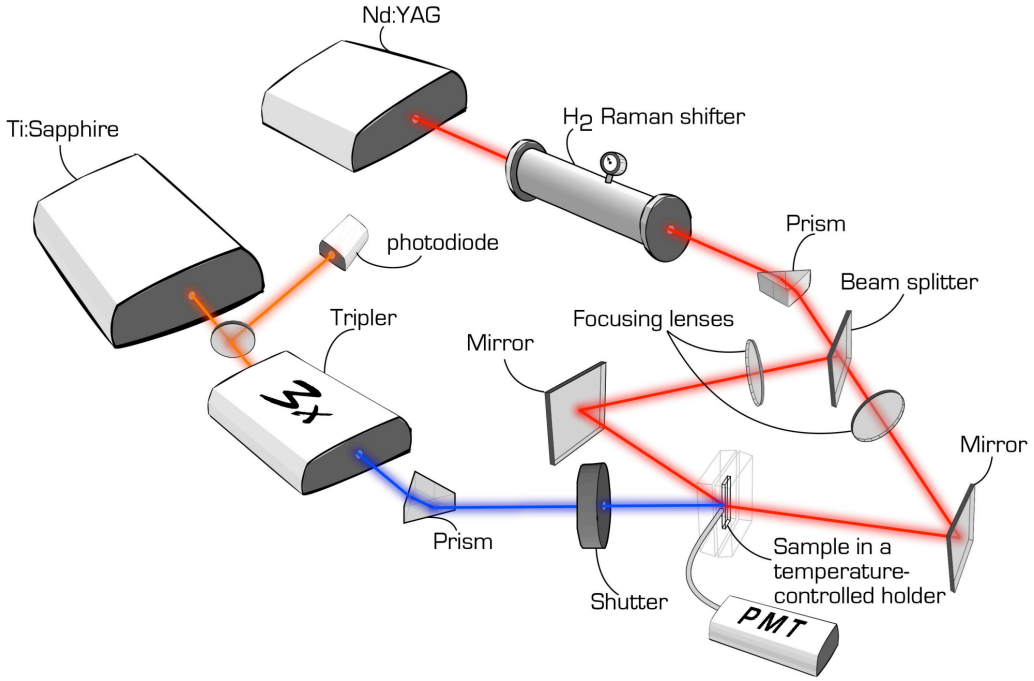


Figure 3.1: Optical components of the nanosecond laser temperature jump instrument.

3.2.4 Data analysis

Singular value decomposition (SVD) was applied to analyze the data obtained from both fluorescence and CD thermal melts (Figures B.1–B.6, Table B.1). In SVD, a data matrix with the spectrum along the x-axis and

temperature along the y-axis is decomposed into three matrices: orthogonal SVD basis vectors that represent the spectra, singular values that represent the importance of each basis vector to reconstruct the spectra, and a trend matrix that shows how each basis spectrum contributes as a function of temperature. The trends that represented the intensity of fluorescence/CD signals (#1) and their wavelength shift (#2) as a function of temperature were then fitted using a two-state thermodynamic model as described previously.1Kinetics data were analyzed using MATLAB (Mathworks Inc., Natick, MA) and IGOR Pro (Wavemetrics Inc., Lake Oswego, OR). Time traces were first averaged and binned into the intervals of 100 decays, which corresponds to 1.25 μs. The time traces were then fitted to a linear combination of the lifetimes averaged between 153.75 and 28.75 μs before the jump ($\chi = 1$) and lifetimes averaged at the last 125 μs of the time traces ($\chi = 0$).The relative lifetime shift as a function of time, $\chi(t)$, was then obtained. The traces were corrected for the effects of photobleaching and cooling after the temperature jump as shown in Figures B.8 and B.9. The time traces were then fitted to a double exponential function starting with point zero where the jump occurred (Figures B.10–B.12).

3.3 Results

3.3.1 λ_{6-85} mutant stability differs by fluorescence and CD detection

We selected a set of λ-repressor mutants spanning a wide range of stabilities (Table 3.1). Nominal melting temperatures (T_m) were determined from two-state fits with linear baselines (Figure 3.2). We reproduced the circular dichroism (CD) melting temperatures (see the Appendix).[31] New measurements of T_m by fluorescence wavelength shifts, which are better correlated with tryptophan fluorescence detected during T-jumps, yielded a different stability ranking. λD14A is ranked most stable by CD[31] but less stable than λQ33Y and λsQ33Y by wavelength shift (Figure 3.2). Such differences indicate a breakdown of two-state behavior.

The mutants were divided into three categories on the basis of fluorescence-detected thermal stability. λQ33Y and λsQ33Y were the most stable vari-

Table 3.1: λ-repressor nomenclature, mutations from wild type, and melting temperatures (rounded to the nearest °C) as determined by fluorescence wavelength shifts

Mutant name	Mutations	T_m (°C)
λQ33Y	Y22W, Q33Y, G46A, G48A	70 ± 0.5
λsQ33Y	Y22W, Q33Y	70 ± 0.5
λD14A	D14A,Y22W,Q33Y,G46A, G48A	68 ± 0.5
λnQ33Y	Y22W, Q33Y, A37G, A49G	53 ± 0.5
λsA49G	Y22W, A37G, A49G	48 ± 0.5

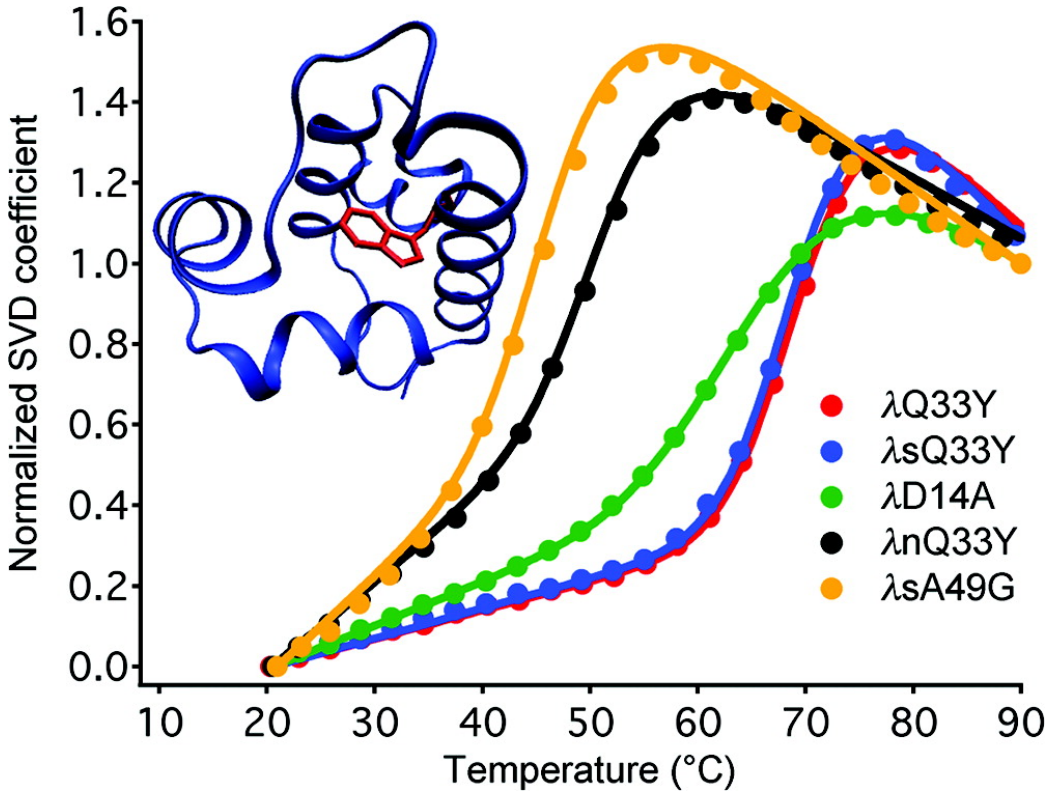


Figure 3.2: Thermal denaturation of the λ-repressor mutants in Table 3.1. The singular value decomposition (SVD) shown here represents the wavelength shift of the W22 fluorescence spectrum (raw data are given in the Appendix). The inset shows W22 in the crystal structure of λQ33Y.[29]

ants, with melting temperatures of 70 °C. The broader melting curve of λD14A occupied an intermediate stability range. λnQ33Y and λsA49G mutants were the least stable of the studied proteins.

3.3.2 Only the two most stable λ-repressor fragments have a significant slow-phase amplitude

We studied the slow relaxation phase of λ-repressor variants by collecting 5 ms of tryptophan fluorescence decays after the T-jump. The tryptophan decay lifetime is sensitive to folding. Individual decays with 3 ns lifetime were sampled at 100 ps intervals for 12.5 ns per decay, yielding a total of 400,000 decays per kinetic trace. Each jump was then reproduced 60 times. Figure 3.3 shows the combined results for three of the mutants binned into 1.25 μs intervals. The function $\chi(t)$ normalizes the tryptophan fluorescence decay time as a folding probe from $\chi = 1$ (shorter lifetime for all Y33 mutants) to $\chi = 0$ (longer lifetime for all Y33 mutants). All data at t > 1 μs were fitted using a double-exponential function with the baseline fixed at zero. With that constraint and the signal-to-noise ratio, slow phases (<10% of the fast-phase amplitude and slower than a few ms) were not discernible in our experiments. The fast initial phase has been discussed in detail elsewhere.[29, 30, 32] The largest slow phases (*e.g.*, Figure 3.3A) could be fitted by observed relaxation times (τ) of 1.2–1.6 ms, and this range also provided good fits for small-amplitude slow phases (if any) of other mutants (for more information, including correction of baselines for recooling and tryptophan photobleaching, see Appendix B).

λD14A, which was investigated computationally by Bowman *et al.*, showed only a ≤10% slow-phase amplitude, nearly within experimental error (Figure 3.3B and Table 3.2). The more stable (by fluorescence melt) λQ33Y and λsQ33Y had by far the largest A_{slow}/A_{fast} ratio among all of the studied mutants (Figure 3.2A, Table 3.2, and Appendix B). Moreover, the amplitude of the slow phase switched when the temperature was increased. This can be explained by a model with at least two non-native states having different tryptophan lifetimes (see below). The glycine-rich and least stable variants, λsA49G and λnQ33Y, showed ≤10% slow phases within experimental error (Figure 3.3C, Table 3.2, and Appendix B).

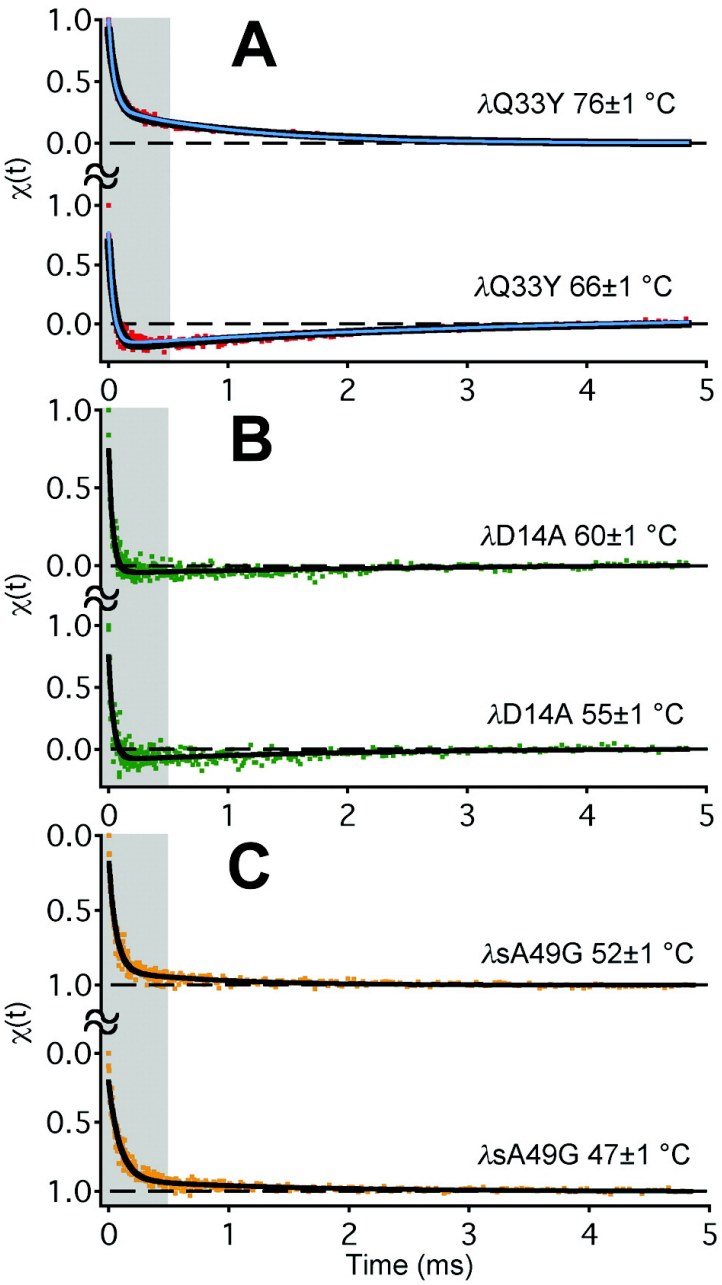


Figure 3.3: T-jump kinetics of λ-repressor mutants detected by tryptophan fluorescence decay, with double-exponential fits in black. The global model fits are shown as blue curves in (A). Traces were normalized from 1 (fast fluorescence decay) to 0 (slow decay). The fast phase (gray areas) was investigated previously,[30–32] so the ms scale is emphasized here.

Table 3.2: Ratios of amplitudes of slow- and fast-folding phases in λ-repressor mutants having different stabilities.

Mutant name	T ($\pm$1 °C)	A_{slow}/A_{fast}
λQ33Y	76	0.37 $\pm$ 0.02
	66	-0.27 $\pm$ 0.03
λD14A	60	-0.08 $\pm$ 0.08
	55	-0.11 $\pm$ 0.09
λsA49G	52	0.1 $\pm$ 0.4
	47	0.1 $\pm$ 0.2

3.3.3 Denaturant produces a slow phase in the D14A mutant

Bowman *et al.* proposed that the simulation might mimic mild denaturant conditions. T-jump experiments were also conducted at low concentrations of guanidine hydrochloride (GuHCl), which was previously reported to induce the formation of the PP/β structure in λ-repressor fragments.[223] Our kinetics experiments showed that 0.5 M GuHCl destabilizes the λ_{6-85} series by 10 °C. The range of the A_{slow}/A_{fast} ratio did not increase in GuHCl for the most stable mutants; they are "maxed out". However, λD14A showed an increased range of A_{slow}/A_{fast} in 1 M GuHCl, beginning to resemble the more stable mutants (Figure 3.4). The least stable mutant, λsA49G, does not completely fold at room temperature in just 0.5 M GuHCl (see Appendix B). Thus, addition of the denaturant does produce slow kinetics for the mutants of intermediate stability, consistent with the suggestion by Bowman et al. that force-field error in the all-atom simulations may correspond to experimental conditions in mild denaturant.

The small millisecond phase of λD14A relative to λQ33Y and λsQ33Y is a genuine indicator that slow dynamics is less important in λD14A and not simply a result of tryptophan fluorescence sometimes being insensitive to slow phases. There is strong evidence that W22 fluorescence can monitor the slow phase whenever it is present in these three mutants. Oas and co-workers suggested that the Q33Y mutation substantially increases the quenching of tryptophan upon folding (private communication), and the Y33–W22 interaction in the native state has been verified by X-ray crystallography of the λQ33Y mutant.[29] The Y33–W22 interaction is equally present in all three mutants on the basis of the large fluorescence changes observed upon melting (Figure 3.2 and Appendix B).

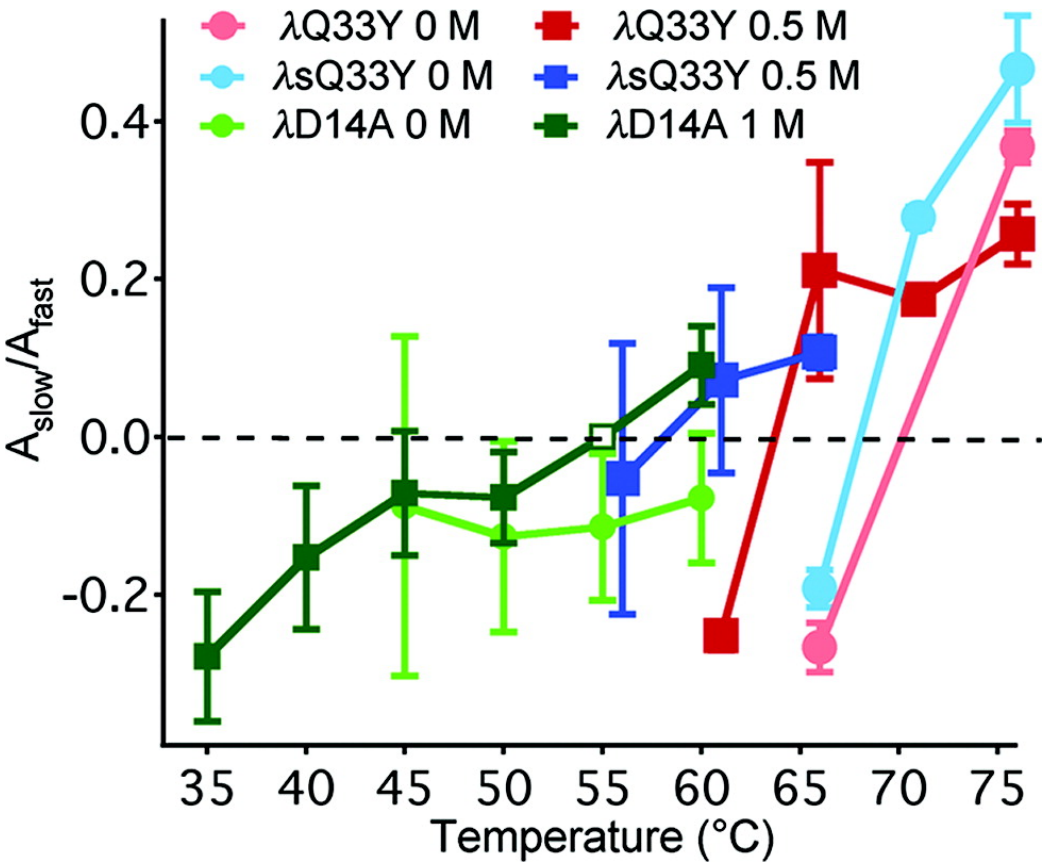


Figure 3.4: Dependence of the A_{slow}/A_{fast} ratio on temperature with and without GuHCl. (open square: fixed point.)

3.4 Discussion and conclusion

3.4.1 A low-resolution landscape model

The unusual result is that the most highly stabilized forms of λ_{6-85} are most likely to get stuck in β-sheet traps, assuming that our slow experimental phase has the same origin as the $\leq$1 ms phases simulated by Bowman *et al.* With two very different time scales (10 μs vs 1 ms for the most stable mutants), our data can be explained by a low-resolution model with three or more states. We will focus on a simple global model consistent with all of the experimental and computational observations to date. The model suggests a specific new hypothesis for future testing.

λ_{6-85} with oxidized methionines has been used to mimic the denatured state under native conditions.[226] It exhibits strong helical propensity for residues 6–20 (centered on helix 1) and reduced flexibility for residues 50–83 (centered on helix 4) but no evidence for a stable PP/β structure. Thus, the unfolded λ-repressor does not necessarily form the compact sheet structure proposed by simulation under native solvent conditions. On the other hand, we recently reported several new mini-proteins containing only λ-repressor helices 1 and 4 connected by a linker.[225] One (λ_{blue1}) has a CD spectrum, denaturation curve, and T-jump kinetics consistent with a cooperatively folding two-helix bundle. The slightly different ?blue3 sequence has a CD spectrum peaked at 215 nm, consistent with β-sheet content.[225] We also observed a CD signature consistent with a PP/β structure in several λ-repressor fragments subjected to combinations of high temperature and mild denaturant.[223] Thus, the PP/β structure is not much higher in free energy than the native α-helical secondary structure.

We fitted the data in Figure 3.3A with several three-state models (see Appendix B). The free energy landscape most consistent with experiment (observed tryptophan lifetime changes) and thermodynamic expectations (extensively unfolded states should have lower free energy than compact states at high temperature) is shown in Figure 3.5A and depicted as a kinetic scheme in Figure 3.5B. The NET model is consistent with both experiment and simulation. The fits to the global model (blue curves in Figure 3.3A) closely match the experimental kinetics. The computed structure we suggest as representative for E in Figure 3.5A has a 0.99 computed probability of reaching

N before reaching T.[33] This is consistent with a rate ratio of 100:1 or a relaxation time ratio of 10 μs:1 ms, close to experimental observation.

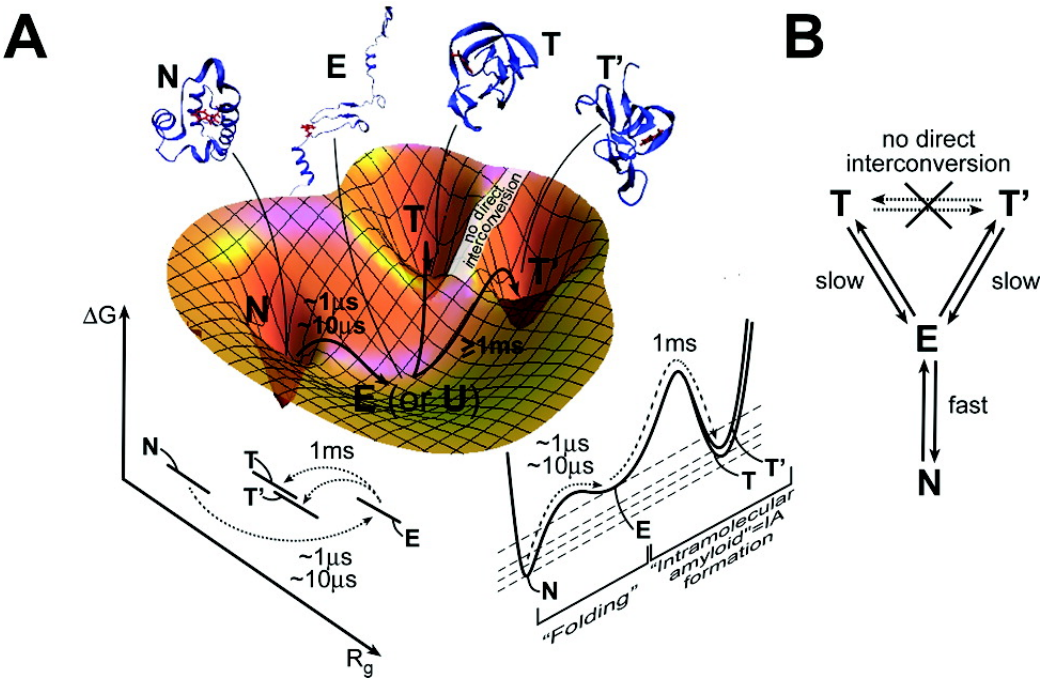


Figure 3.5: Folding of λ_{6-85} mutants along reaction coordinates related to compactness (R_g) and non-native secondary structure content. (A) Free energy landscape. The folding process per se from an extensively unfolded state E (or U) to N is fast. E can also convert slowly to compact non-native states T and T′. Suggested structures are from refs [33] and [29] (N). (B) Kinetic scheme. T and T′ interconvert indirectly through E, which is in rapid equilibrium with N.

In the NET model, the extensively unfolded state E interconverts in microseconds with the native α-helical state N. This process is what one ordinarily considers as protein folding. In addition, E interconverts in milliseconds with traps such as T or T′. We fitted such traps as a single kinetic state because our fluorescence probe cannot distinguish them.

Kinetic traps of various types are known for many proteins,[227] but the simulations of Bowman *et al.* make very specific structural predictions in addition to predicting a millisecond relaxation time: different traps T and T′ resemble each other in compactness and high β-sheet content but differ in the detailed arrangement of the sheets. The slow interconversion of T and T′ is thus explained because their non-native secondary structure has to be unmade via the extensively unfolded state E. Once E is reached, it

very rapidly interconverts with N. Therefore, N is sampled many times on the time scale of T and T′ interconversion. In our view, the native state N is thus not so much a "hub" but is simply easily reached from E, if we distinguish the N⇔E "folding process proper" from the E⇔T,T′ misfolding process.

3.4.2 Hypothesis

We suggest the following hypothesis to explain the structures seen by Bowman *et al.* as well as the slow phase observed experimentally only for the most stable λ_{6-85} mutants: Proteins whose hydrophobic interactions strongly favor compact states are prone to form compact β-rich traps or "intramolecular amyloids" (IAs). There is evidence for a PP/β structure in monomeric denatured states of many proteins.[223, 228] Such non-native secondary structure could get locked in when proteins rapidly become compact during folding, especially under mildly denaturing conditions. In particular, large proteins with many nonlocal options for hydrophobic contacts and β-strand formation might be more likely to form IAs in need of rescue by chaperones. Direct structural evidence for IAs could come from two-dimensional IR kinetics studies. IAs are complementary to a phenomenon observed for some proteins with a β-rich native state: formation of non-native helical traps facilitated by favorable local interactions.[229]

Proteins have some kinetic and thermodynamic protection from IAs: Kinetically, the extensively unfolded state E converts much more rapidly to N than to T, in analogy to the large barrier separating real amyloid aggregates from denatured proteins and native states. Thermodynamically, T still ends up higher in free energy than N (unlike real amyloid aggregates; see the model fit in Appendix B). From an energy landscape perspective, one would say that fast folding λ-repressors have a glass transition temperature T_g dangerously close to the folding temperature T_f[230] but still end up getting stuck in traps only temporarily.

CHAPTER 4

MISPLACED HELIX SLOWS DOWN ULTRAFAST PRESSURE-JUMP PROTEIN FOLDING

4.1 Introduction*

Temperature and pressure are excellent perturbations when comparing experimental folding kinetics with molecular dynamics (MD) simulations [231]. Fast temperature-jumps (T-jumps) and pressure-jumps (P-jumps) are relatively easy to simulate by MD. Typical temperature changes required to cross the protein folding transition are 5–20 K, easily implemented with laser T-jumps [232]. Typical pressure changes required to cross the folding transition are 1–5 kbar, previously achieved only with millisecond time resolution (piezo methods are limited to $\Delta P < 100$ bar) [233, 234]. We recently reported a P-jump instrument capable of >1-kbar P-drops with ∼1-μs dead time [235].

Folded proteins have a larger partial molar volume than pressure-denatured proteins (by about 10^1–10^2 mL/mol) [236]. The fractal dimension of their folded state is less than 3 because voids occur whenever a connected chain made from a finite amino acid alphabet is packed into a compact structure [237, 238]. Such imperfections in packing, which disappear when small water molecules solvate the polypeptide chain, lead to protein unfolding under pressure [239]. Pressure unfolding is a slow process because the positive activation volume is unfavorable at high pressure [240].

Here, we study the much faster process of protein refolding at 1 bar and room temperature, starting from the pressure-denatured ensemble. We chose λ*YA, the Y22W/Q33Y/G46,48A mutant of λ-repressor fragment 6–85, as our model protein [241, 242]. Tryptophan W22 provides a fluorescent probe [241]. Based on the crystal structure, tyrosine Y33 enhances the fluorescence

* This chapter is partially reproduced from Maxim B. Prigozhin, Yanxin Liu, Anna Jean Wirth, Shobhna Kapoor, Roland Winter, Klaus Schulten and Martin Gruebele, Proceedings of the National Academy of Sciences of the United States of America, 110(20), 8087–8092, 2011, DOI: 10.1073/pnas.1219163110

lifetime difference between folded and unfolded states by contact quenching W22 when the protein is folded [34]. The glycines-to-alanines substitution stabilizes helix 3 of the protein (Figure 4.1); thus, helices 1, 3, and 4 have high stability [243]. Stable helical structure in small peptides is difficult to pressure-denature [244], even though secondary structure, whose stability is assisted by tertiary contacts, denatures easily when the tertiary contacts are disrupted [245]. With the high stability of helices 1, 3, and 4 in λ^*YA, we expect a relatively helix-rich unfolded coil on P-denaturation of λ^*YA, even in the 2.4 M guanidine hydrochloride (GuHCl) buffer we use to poise the protein at the unfolding transition.

We find that pressure-denatured λ^*YA is very different from temperature-denatured λ^*YA, offering an opportunity to study the effect of the initial denatured state on refolding. Pressure unfolding induces a mix of helix and coil conformations, and it is reversible at all pressures used in our high-pressure experiments. In contrast, high temperature causes aggregation or populates extended (β-like) structure at high denaturant concentration [246]. Our microsecond resolution P-jump refolding experiment reveals a 1.4-ms "slow" phase, in addition to a microsecond "burst" phase. Based on residual helical structure in the pressure-denatured state as measured by IR spectroscopy, we assign the slowdown to a trap with nonnative helix, which does not exist in the temperature-denatured state.

To investigate the plausibility of this assignment further, we calculated over 50 μs of explicit solvent MD trajectories using four different force fields. First, λ^*YA was denatured at high temperature and pressure *in silico*. With CHARMM27 and 22* force fields, refolding trajectories after a downward P-jump reveal a likely culprit for the slow phase: misplaced helical structure where there should be loops connecting helices in the native state. Such nonnative helix forms rapidly *via* local interactions and then traps the protein in compact states because incorrect loops lead to incorrect packing of the native-like helices. By identifying loop regions with nonnative helix propensity, we predict mutation sites that could reduce the slow recovery from the misplaced helix traps. CHARMM36 and AMBER99-SB do not rapidly form compact helix-rich states after the P-jump. The initial state before the jump is not to blame: CHARMM27 and 22* form helical traps even when they start out with initial conditions from AMBER-99SB.

Nonnative transient helix en route to the native state has been observed

many times under cryogenic conditions [247–249], but it can be studied in our experiments at room temperature and 1 bar. Although helix overshoots in MD simulations may be due to force field bias [250], our results highlight that misplaced helical structure may occur experimentally also. Microsecond P-jump studies can contribute to the continued calibration and improvement of force fields by providing an alternative initial ensemble to temperature denaturation.

4.2 Methods

4.2.1 Protein expression and purification

λ^*YA was expressed as described previously [251]. GuHCl, deuterium oxide 99.9% (D_2O), and NATA were obtained from Sigma-Aldrich and used without further purification.

4.2.2 P-jumps

P-jumps with a home-built apparatus are described in detail elsewhere [235]. Briefly, 300 μM protein solution was pipetted into an $\sim$8-μL hemispherical dimple machined in-house into an optical grade 3/8-inch sapphire cube (Esco Products). The dimple was then covered with a double-layer of mylar-coated aluminum foil to prevent mixing of the sample with ethanol, which served as the pressurization fluid. The cube was inserted into a pressurization clamp as described elsewhere [235], and the sample was pressurized to 1.2 kbar using a hydrostatic pump (High Pressure Equipment Company). We used 0.007-inch-thick stainless-steel burst membranes and 95 V ($\sim$10-kA current) to burst the membrane.

The sample was excited by a Ti:sapphire laser (KMLabs). The laser emission at 840–860 nm was frequency-tripled to 285 $\pm$ 3 nm. The laser was adjusted to a pulse rate of 80 MHz. Tryptophan fluorescence decay was collected by means of an optical waveguide (Oriel) every 12.5 ns and passed through a B370 band-pass filter (Hoya) onto a photomultiplier (R7400U-03; Hamamatsu Corp.). The fluorescence decays were recorded by a DPO7254 digitizer (Tektronix) with 2.5-GHz bandwidth, locked to the 80-MHz laser

cavity to avoid aliasing. Each kinetic trace contained 400,000 fluorescence decays. Each fluorescence decay was sampled at 10 GHz (125 points per decay). The amplitude of the fluorescence signal was 100–250 mV.

The data were analyzed using custom code written in LabView (National Instruments), MATLAB (MathWorks), and IGOR Pro (Wavemetrics). One hundred raw fluorescence decays were binned into average decays every 1.25 microseconds, reducing the time resolution to a value comparable to the dead time of the P-jump. To assign a single lifetime parameter to each fluorescence decay, we fitted them to the linear combination of a prejump and postjump fluorescence decay. We averaged the first 100 decays of the time trace to get a representative decay f_1 before the P-jump and the last 100 decays at 5 ms to get a representative decay f_2 after the P-jump. Decays between these two P-jumps were fitted by $f = a_1 f_1 + a_2 f_2$, and the fraction contributed by the second decay was calculated as $\chi = a_2/(a_1 + a_2)$. Thus, a decay with the same lifetime as before the jump yields $\chi = 0$, and a decay with the same lifetime as at 5 ms yields $\chi = 1$. We fit the resulting time trajectory to a double-exponential function $A_f e^{-t/\tau_f} + A_s e^{-t/\tau_s}$.

4.2.3 Fluorescence thermodynamics under pressure

Fluorescence spectra were measured using a Cary Eclipse fluorescence spectrophotometer (Varian). Excitation and emission slit widths were 5 nm each, excitation wavelength was 280 nm, and the scan rate was 120 nm/min. Sample concentration was 200 μM. The sample was pressurized with a high-pressure cell (ISS). We used a rectangular quartz cuvette with a path length of 4 mm. Spectrophotometric grade ethyl alcohol (95.0%, A.C.S. reagent; Acros Organics) was used as pressurization fluid. The center of the spectral mass in Figure 4.1B is the weighted average $\langle I \rangle = \frac{\int d\lambda\ \lambda\ I(\lambda)}{\int d\lambda\ I(\lambda)}$ of the fluorescence intensity in Figure 4.1A [236].

4.2.4 IR thermodynamics under pressure

Lyophilized protein was dissolved in 50 mM K_3PO_4 buffer in D_2O (Sigma), and D_2O exchange was allowed to proceed for 2 h at room temperature and then overnight at 4 °C. The protein was then lyophilized again. For the

measurement, the protein was dissolved in 50 mM K_3PO_4 buffer in D_2O at pD = 7.1 to the final concentration of 1.7 mM. pH of deuterium (pD) was adjusted using DCl. Approximately 10 μL of the sample was then placed in a diamond anvil cell equipped with type IIa diamonds (High Pressure Diamond Optics). The sample was held in a stainless-steel gasket (overall diameter = 12.5 mm, center pinhole diameter = 0.45 mm, thickness = 0.050 mm). The gasket was secured in the diamond anvil cell using gum from Faber–Castell (127020). A small amount of $BaSO_4$ was placed in the chamber with the sample. IR spectra were collected using a Nicolet Magna IR 550 spectrometer equipped with a nitrogen-cooled mercury cadmium telluride (MCT) detector. The sample chamber within the instrument was purged with dry CO_2-free air. The IR beam was focused onto the pinhole of the diamond anvil cell. Spectra were collected from 400 to 4,000 cm^{-1} using 256 accumulations and a resolution of 1 cm^{-1}. Pressure was increased using the spring-loaded screw of the diamond anvil cell assembly. Changes in the pressure were quantified by monitoring the stretching vibration of the calibrant, $BaSO_4$. Its peak was at 983.6 cm^{-1} at 1 bar and shifted linearly with pressure toward larger wave numbers.

4.2.5 All-atom MD simulations

All-atom MD simulations were performed in explicit solvent using the TIP3P water model [252]. The simulations were carried out both on general-purpose supercomputers using NAMD 2.8 [253] and on the special-purpose supercomputer Anton [254]. All simulations were carried out with periodic boundary conditions in constant particle number, temperature and pressure ensemble. More details of the simulations are given in Appendix C.

To generate two initial states for P-drop simulations, we started with λ^*YA constrained to the PDB structure (PDB ID code 1LMB) [255] at 325 K. The pressure was increased from 1 to 5 kbar in 0.15 μs. The temperature was then ramped to 525 K for another 0.15 μs to unfold the protein more extensively (Appendix C, Figure C.8). Finally, the simulation was returned to 325 K and 5 kbar to yield the initial states for the P-drop simulations.

4.3 Results

We used λ^*YA as a model system to study the behavior of a fast folder by P-denaturation and microsecond P-jump kinetics. We used two different probes in our experiments, fluorescence of tryptophan W22 and IR absorbance of the peptide bond in the amide I′ region of the IR spectrum. W22 fluorescence was used previously to study fast folding of λ^*YA by T-jumps [34, 242]. Thus, our pressure experiments can be directly compared with the T-jump results. The motivation to use IR absorbance as a complementary structural probe came from recent computational and experimental reports indicating the presence of transient β-sheet formation in λ^*YA at high temperature [246, 256, 257]. It was unknown what type of residual secondary structure would exist in λ^*YA at high pressure.

4.3.1 Equilibrium pressure denaturation probed by fluorescence

λ^*YA was prepared just shy of denaturation in 2.4 M GuHCl (Appendix C, Figure C.2). Subsequent pressure denaturation of λ^*YA was monitored by tryptophan fluorescence spectra and fluorescence decay lifetimes of 200–300 μM λ^*YA in 50 mM K_3PO_4 buffer (Methods and Figure 4.1). Decay lifetime (and also intensity) is sensitive to tryptophan quenching by tyrosine 33, which was used as the probe in our P-jump. The peak shift of the spectra is sensitive to solvent exposure (redder = more exposed).

Without denaturant, the small shift of the spectrum to longer wavelengths (Figure 4.1B) is very similar to that observed for the tryptophan derivative N-acetyl-tryptophanamide (NATA). Like NATAs lifetime, the lifetime of λ^*YA decreases linearly with pressure in the absence of denaturant (Figure 4.1D). Pressure denaturation of λ^*YA in 2.4 M GuHCl (Figure 4.1) shows a much larger wavelength shift and a nonlinear lifetime increase. By fitting the fluorescence spectra in Figure 4.1A, we estimate ≈40% denatured population at 1,200 bar, the starting point for our P-jump experiments. It is common for pressure-induced protein denaturation to have low cooperativity [235, 258–260], and this is also the case for λ^*YA in Figure 4.1.

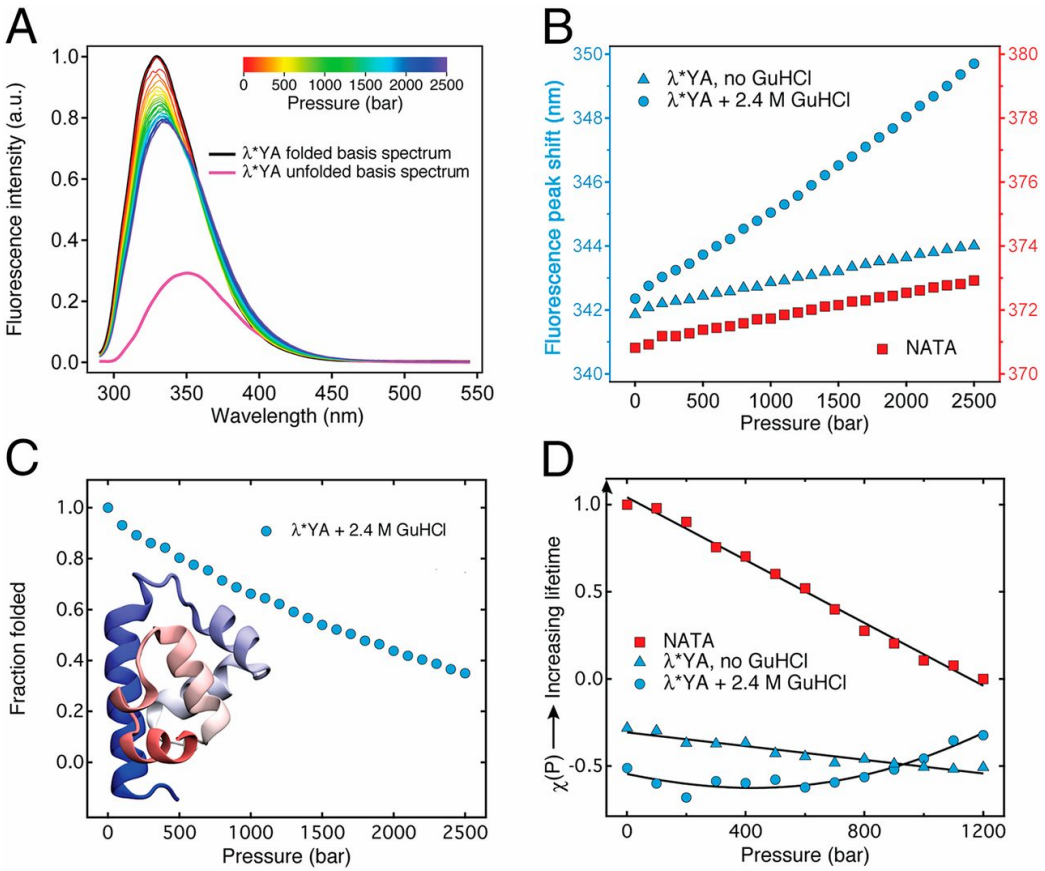


Figure 4.1: Pressure denaturation of λ*YA probed by fluorescence spectroscopy of a 200-μM sample (A–C) and fluorescence lifetime analysis of a 300-μM sample (D). (A) Fluorescence spectra of λ*YA in guanidine (*p*H 7) at 100-bar intervals from 1 to 2,500 bar (rainbow gradient). The basis spectrum of the folded state is shown in black, and that of the fully unfolded state is shown in purple (Appendix C). a.u., arbitrary units. (B) Fluorescence peak shift (centroid) as a function of pressure. λ*YA in 2.4 M GuHCl shows a much larger 0.0032-nm/bar shift than λ*YA in buffer (native state model) or NATA (denatured state model). (C) Fraction folded was calculated by fitting the spectra in A to a linear combination of the folded and unfolded basis spectrum (two-state model; Appendix C); at 1,200 bar (initial condition for P-jumps), ≈40% of the protein is unfolded. The crystallographic structure of λ*YA obtained from the PDB (ID code 3KZ3) is shown. (D) Scaled fluorescence lifetime change relative to NATA (1 at 1 bar, 0 at 1,200 bar). NATA and λ*YA in 0 M GuHCl lifetimes decrease linearly with pressure, whereas λ*YA in 2.4 M GuHCl shows the onset of pressure denaturation (χ for proteins was shifted up by +3 because NATA has a much longer lifetime; Appendix C, Figure C.4).

4.3.2 Fast P-jump kinetics

We induced relaxation of λ^*YA from the denatured state to the native state by means of a large microsecond P-drop. We extended the capabilities of our recently reported P-jump instrument [235] further: Mechanical damping and optical isolation make the triggering more reliable, and up to 5 ms of data can be collected (0.5 ms previously). The protein solution is placed into a 1-mm dimple machined into a sapphire cube, which is optically transparent at wavelengths greater than 280 nm for laser excitation and fluorescence detection (Figure 4.2). We monitored pressure denaturation of λ^*YA via tryptophan lifetime (Figure 4.1D) by slowly increasing the pressure of the sample to 1,200 bar.

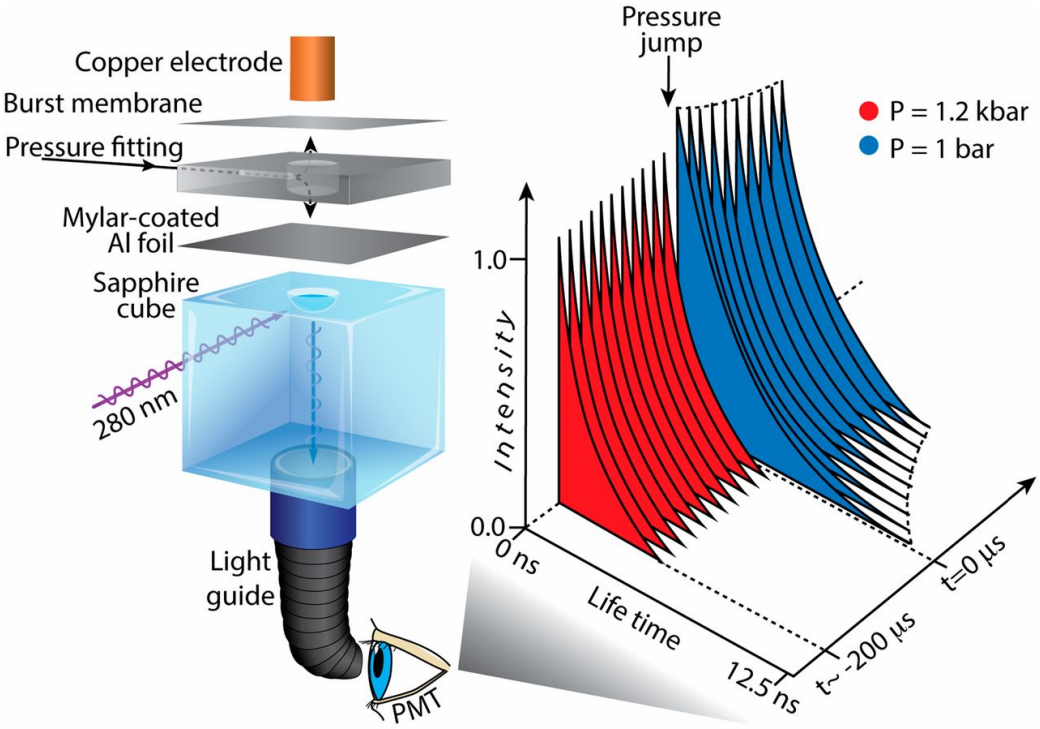


Figure 4.2: P-jump instrument. The sample is pipetted into a dimple in a sapphire cube. The dimple is covered with mylar-coated aluminum foil and pressurized by pumping ethanol into a pressure fitting. A current burst into a copper electrode bursts the upper steel membrane and releases the pressure. Sample fluorescence is excited by a 280-nm pulsed laser every 12.5 ns and is collimated by a UV light guide onto a photomultiplier. The digitized raw data consist of a train of fluorescence decays, whose lifetime and intensity monitor the refolding of the sample after the sudden P-drop at $t = 0$.

The pressure was then jumped down to 1 bar with a microsecond dead time by resistively heating and puncturing a steel burst membrane (Figure 4.2). To monitor kinetics, the sample was excited by 280-nm UV pulses every 12.5 ns. Tryptophan fluorescence decays were detected every 12.5 ns with a 100-ps time resolution. To monitor how the fluorescence lifetime relaxes to equilibrium, we applied linear two-state fitting to the decays to scale the lifetime change from $\chi = 0$ (before P-jump) to $\chi = 1$ (5 ms after P-jump) (Methods). The <3-μs dead time of the P-jump instrument was calibrated by performing pressure jumps on NATA (Appendix C, Figure C.1).

Figure 4.3 shows the refolding kinetics of λ^*YA on a 1,200→1-bar P-jump at 295 K in buffered 2.4 M GuHCl. Two phases are observed: a fast microsecond phase during which the W22 lifetime increases relative to the value before the P-jump and a slower millisecond phase during which the W22 lifetime decreases. The fast phase could not be resolved within the dead time. For reference, we also measured the P-jump kinetics of λ^*YA without denaturant, our model for folded protein (Figure 4.1). The net burst phase is the difference between the blue and red curves at t = 0 (Figure 4.3). On the relative scale of the NATA fluorescence lifetime change (0 at 1,200 bar, 1 at 1 bar), the net amplitude of the burst phase is +0.2 and the amplitude of the ms phase is –0.4 (absolute changes in fluorescence decays are shown in Appendix C, Figure C.4). The fast phase associated with protein refolding has a relaxation time $\tau_f \leq 3$ μs, and the slower phase fits to a single-exponential lifetime of $\tau_s = 1.4 \pm 0.2$ ms. The T-jump unfolding relaxation kinetics from Prigozhin and Gruebele [257] at 339 K are also shown for comparison. A ≈60-μs phase of increasing lifetime was observed in that case, followed by a small millisecond phase attributed to a β-sheet-rich trap in the studies by Bowman *et al.* [256] and Prigozhin and Gruebele [257].

These observations are consistent with two scenarios. In one scenario, separate microsecond- and millisecond-folding populations start out from two slowly interconverting pressure-denatured states and proceed to the native state, where W22 is quenched by Y33. Slow interconversion is not implausible at high pressure, given the potentially large activation volume between compact denatured states. In the other scenario, the entire denatured protein population is trapped in a partially folded state within 3 μs, from which it escapes to the native state within 1.4 ms. In both scenarios, it is also possible that transient aggregation contributes to the slow millisecond phase.

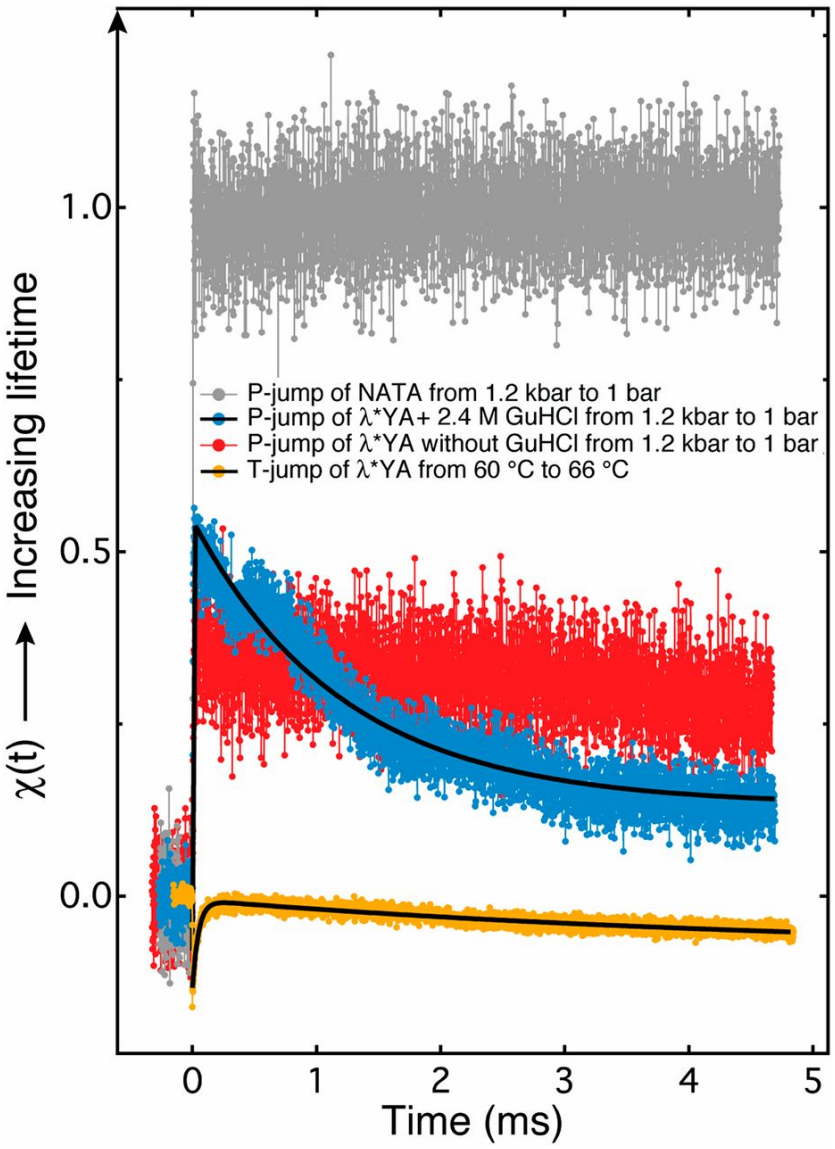


Figure 4.3: P-jumps (300-μM sample) and T-jumps (200-μM sample) of λ^*YA and NATA, probed by tryptophan fluorescence decays. Tryptophan lifetime change was normalized for NATA so that $\chi = 0$ corresponds to the decay lifetime before the jump (1,200 bar) and $\chi(t) = 1$ corresponds to the decay lifetime 5 ms after the jump (1 bar). The rest of the jumps were analyzed using the lifetime decays from the P-jump of NATA for direct comparison. Solid black curves are the double-exponential fits of the data with relaxation times $\tau_f = 3.8 \pm 0.4$ μs and $\tau_s = 1.4 \pm 0.2$ ms for the P-jump and $\tau_f = 63 \pm 2$ μs and $\tau_s = 2.17 \pm 0.02$ ms for the T-jump.

4.3.3 Pressure denaturation probed by IR absorption

To relate the slow phase on P-drop refolding of λ^*YA to secondary structure, we used IR absorption spectroscopy. IR spectroscopy has been successfully applied in the past to resolve the secondary structure content of proteins under high pressure [245]. We carried out equilibrium pressure denaturation of λ^*YA in a diamond anvil cell. We monitored pressure-induced changes in the amide I′ region (prime denotes measurements in D_2O solvent) from 1,600 to 1,700 cm^{-1}. This region contains vibrational information about the carbonyl group in the amide bonds, which depends on the dihedral angles of the backbone, and therefore the secondary structure of the protein. The crystallographic state of λ^*YA is 71% helical [Protein Data Bank (PDB) ID code 3KZ3 [34]]. The IR spectrum of a helix in the amide I′ region contains a single band with a peak wave number at ≈1,650 cm^{-1}. Under pressure, the protein is expected to lose at least some of its helical content in favor of the random coil structure. The spectrum of a random coil exhibits a characteristic maximum at ≈1,640 cm^{-1}.

Figure 4.4A shows IR spectra of λ^*YA in the absence of denaturant, one recorded at 1 bar and the other recorded at 13.9 kbar. The absorption maximum shifts from 1,650 cm^{-1} (helix) toward 1,640 cm^{-1} (coil) on pressure denaturation. The two spectra were used as basis functions to fit a sequence of λ^*YA spectra as a function of pressure. A linear combination of these two basis functions fitted data at all pressures within measurement uncertainty (Appendix C, Figure C.5). Figure 4.4B shows the pressure denaturation curve, along with a two-state thermodynamic fit [251]. The data without denaturant show a transition midpoint at 6.0 ± 0.2 kbar. Helix-coil denaturation is also observed in 0.5 M guanidine deuteriochloride (GuDCl) by IR spectroscopy, and the absence of aggregation can be verified in 2.4 M GuDCl (Appendix C, Figure C.7).

Temperature denaturation produces an entirely different result. The gray curve in Fig. 4.4A is the IR spectrum at 1 bar and 368 K. At 335 K, the protein undergoes a cooperative thermal denaturation. Two shoulders at 1,618 cm^{-1} and 1,680 cm^{-1} appear, due to formation of extended [β-sheet or left-handed polyproline helix (PP-II)] structure at high temperature. Extended structure in temperature/GuHCl-denatured λ-repressor can occur in the monomeric protein [246]. Indeed, at the protein concentration of 2.5 μM,

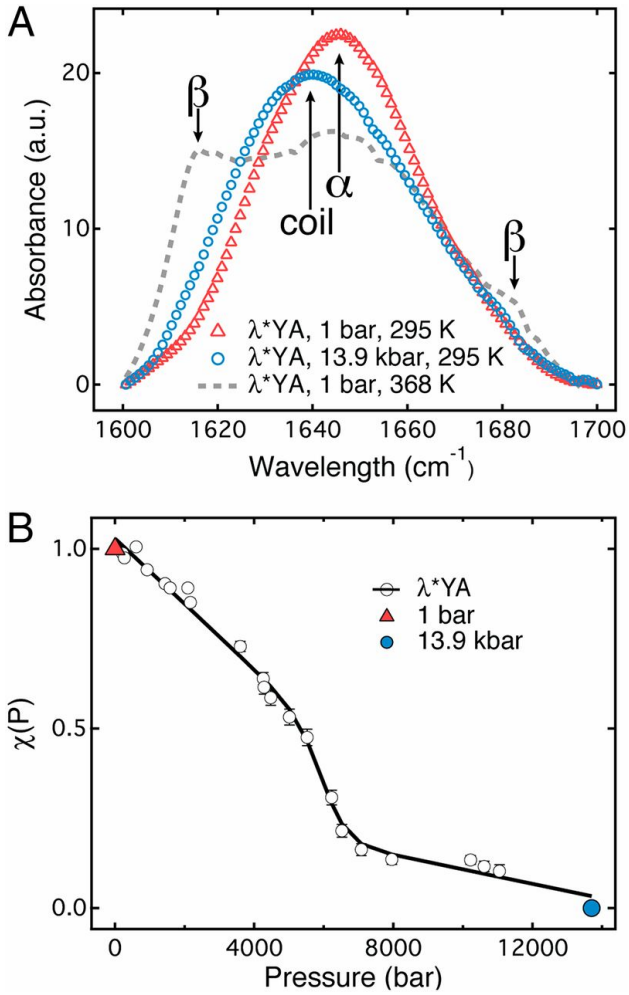


Figure 4.4: Equilibrium denaturation of λ*YA by pressure and temperature, probed by IR spectroscopy (1.7-mM sample). (A) IR absorbance spectra of λ*YA in the amide I′ region measured at 295 K. Triangles indicate 1 bar, and circles indicate 13.9 kbar. These spectra were used as basis functions for the analysis of the entire pressure denaturation curve (Methods and Appendix C). The IR absorbance spectrum of λ*YA in the amide I′ region measured at 1 bar and 368 K is shown as a gray dashed line. (B) Denaturation of λ*YA as a function of pressure [$\chi(P) = 1$ means the 1-bar basis function contributes 100% of the signal, $\chi(P) = 0$ means the 13.9-kbar basis function contributes all the signal]. A thermodynamic two-state fit of the data is shown as a solid black curve, and the error bars are the residuals (Appendix C, Figure C.5). The midpoint of pressure denaturation, P_m, is equal to 6.0 ± 0.2 kbar in the absence of denaturant.

temperature denaturation is reversible (Appendix C, Figure C.3). However, the IR measurement was taken at a much higher concentration (1.7 mM), and above 50 μM, denaturation is irreversible and aggregates form (Appendix C, Figure C.3).

In contrast, Appendix C, Figure C.6 shows full reversibility of pressure denaturation at the protein concentration used for our pressure jumps, and Figure 4.4A shows no evidence of β-sheet structure. Thus, the pressure-denatured state of λ^*YA is very different from the temperature-denatured state. The denatured state preceding the downward P-jump in Figure 4.3 is a helix-coil state, rather than a β-rich state as proposed for high-temperature kinetics [256, 257]. Without any evidence at all of equilibrium pressure-dependent aggregation with or without denaturant (Appendix C, Figure C.6), we also think that transient aggregation is unlikely to dominate the slow 1.4-ms phase in the two scenarios described above, but we cannot rule out transient interactions of the helix-coil states during folding.

4.3.4 P-denaturation and P-jump refolding probed by MD simulation

To probe fast protein refolding on downward P-jump with atomistic detail, we carried out all-atom MD simulations in explicit solvent with four force fields: CHARMM27, CHARMM22*, CHARMM36, and AMBER99-SB [261–266] (Methods).

Initial denatured structures at 325 K and 5 kbar were created by a 0.3-μs heating/pressurization protocol (Methods, Appendix C and Figure C.8). The protocol differs in three important aspects from experiment. No denaturant is included in the simulation. MD pressure unfolding was assisted by high temperature because pressure unfolding alone is very slow. Finally, refolding was studied at 325 K instead of 295 K because the melting point of λ-repressor is too high in CHARMM27 and CHARMM22*; thus, 325 K in silico correlates with a lower experimental temperature [264].

Two fast pressure drops were simulated with CHARMM27 (Figure 4.5). Analogous results for CHARMM36 and AMBER99-SB are shown in Appendix C, Figure C.10. The two refolding trajectories in Figure 4.5 collapse rapidly into structures with near-native radii of gyration, making occasional

excursions to a larger size in the search for the native state. Both accumulate nonnative helical structure within 2 μs after the P-jump (Appendix C, Figure C.9). The first trajectory yields a compact helical globule. The second trajectory produces a conformation much closer to the native state (Appendix C), with helices 1–3 formed and packed into the correct native orientation (Appendix C and Figure C.9). However, helix 4 is threaded through the ring formed by helices 1–3, creating a knotted conformation (Figure 4.5, Upper Right).

In Figure 4.6, the α-helical propensity for each residue is given based on the time percentage spent in α-helical conformation in the last 8 μs of each simulation. Both trajectories form nonnative helix, where there should be turns or loops in the native state (red arrows in Figure 4.6). A compact trap with helix misplaced into loops that persists for >10 μs is consistent with the experimental data of a slowly (1.4 ms) refolding state formed from a helix-coil-rich denatured state right after the P-jump.

4.3.5 MD simulations using different force fields

Protein dynamics in MD simulations depends on the underlying force field. CHARMM27 with correction map [263] has a bias toward helical structure (*e.g.*, ref. [250]). In contrast, our CHARMM36 and AMBER-99SB simulations with the exact same protocol produce virtually no helix in the initial denatured structure (Appendix C, Figure C.8), nor do they produce any native-like structure or a helical trap on P-drop (Appendix C, Figure C.10).

We examined if the initial configuration of the protein is responsible for inability of helical structure to form rapidly. We started with the AMBER-99SB denatured structures from Appendix C, Figure C.8 and then simulated the P-drop with CHARMM27 and CHARMM22* force fields at T = 325 K and P = 1 bar (Appendix C, Figure C.11). The CHARMM22* force field has both folded helical-rich protein and β-rich protein [267]. Although there is almost no helix left in the initial configuration, the helix recovered to near-native value in less than 1 μs with CHARMM27. In a 10-μs simulation using CHARMM22*, helix content becomes significant, although it never exceeds the native value. Like CHARMM27, CHARMM22* exhibits a nonnative helical propensity in several turn/coil residues (M42, G43, A56, L57, and

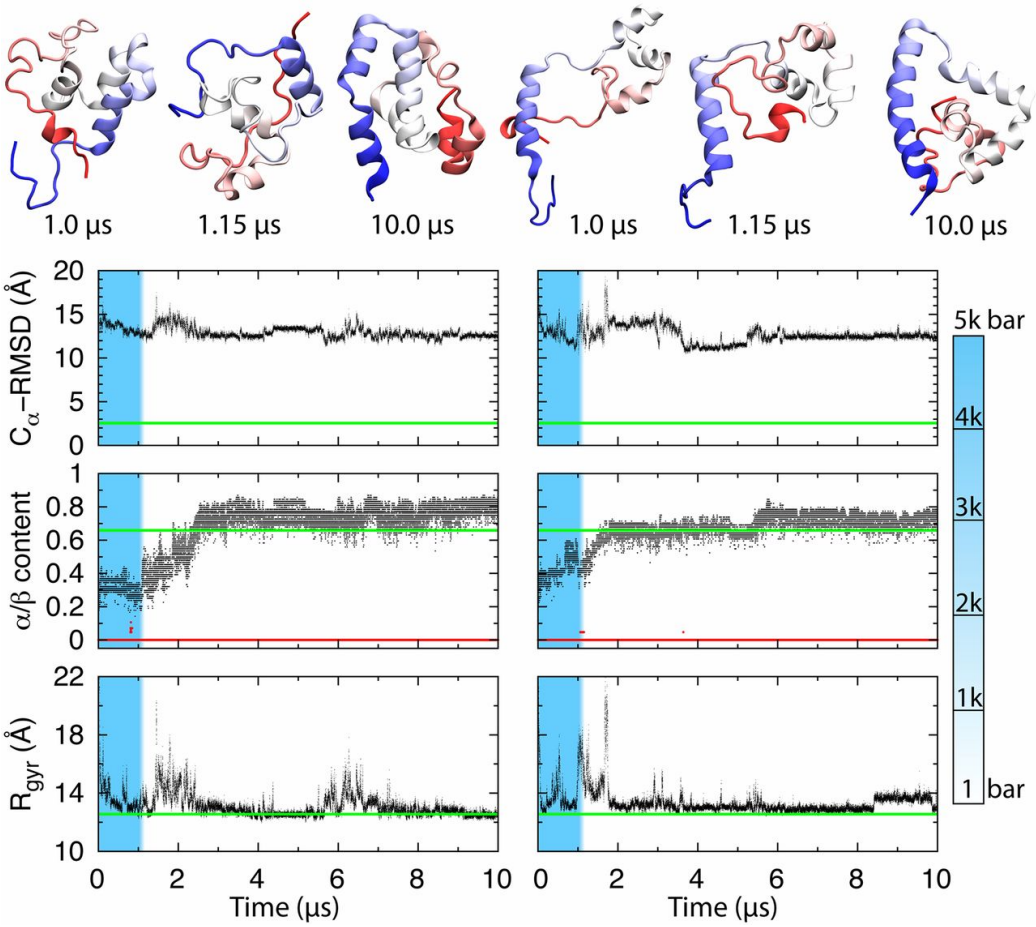


Figure 4.5: CHARMM27 simulation of λ^*YA during P-drop. (Upper) Structures from the two trajectories. The high-pressure simulations start with 1 μs at 325 K and 5 kbar (blue zone), followed by a 0.15-μs P-drop to 1 bar (white zone). (Lower) Refolding (8.85 μs) was simulated at 1 bar and 325 K. Central carbon atom root mean square displacement values were calculated relative to the crystal structure (PDB ID code 3KZ3) [34]. The fraction of residues in α (gray) and β (red) conformations is shown. R_{gyr} is the unsolvated radius of gyration. The native mean values (green solid lines, except red for β-fraction) are from a 150-ns equilibrium simulation of the native structure at T = 325 K and P = 1 bar.

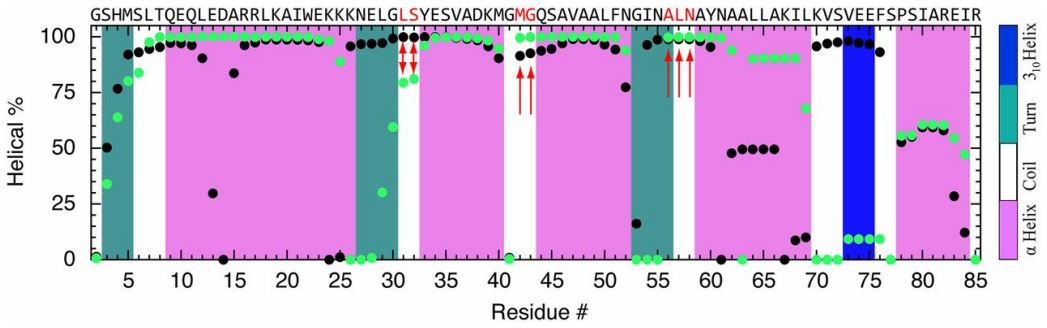


Figure 4.6: Residue-specific α-helical propensity of the simulations in Figure 4.5 (black, first simulation; green, second simulation). The helical percentage was defined as the time percentage each residue spent in α-helical conformation during the last 8 μs of refolding simulation. The secondary structure of the crystal structure is shown as a color-coded background, and the sequence at the top, together with the red arrows, highlights turn/coil residues with >75% helix content in both simulations.

N58 in Appendix C, Figure C.12).

CHARMM22* and CHARMM27 are thus most consistent with our experimental data. Although there is a strong helical bias in the CHARMM27 force field, another variant of λ-repressor was folded successfully in a high-temperature enhanced sampling MD simulation using CMARMM27 force field [268]. The high temperature used in the current study likely compensates for the helical bias in the force field to match up best with the experiment.

4.4 Discussion

Fast-folding experiments and simulation have converged to the point where several small proteins folding equilibria have been observed in silico and validated by kinetic, structural, and thermodynamic experimental data [267–269]. The same small model proteins, under the right conditions, now offer the opportunity to study delays in folding caused by a frustrated free energy landscape or transient aggregation. The advantage of these model proteins is that their misfolding processes can also be quite rapid, and are thus amenable to simulation [256].

The formation of helical secondary structure requires only local i to $i + 4$ contacts. Thus, it is not surprising that helical intermediates that trap pro-

teins en route to the native state have been reported for a wide variety of proteins [247–249]. However, these excess helix traps have generally been observed under cryogenic conditions.

Here, we provide experimental and computational evidence that the five-helix bundle λ^*YA can get trapped transiently in states with misplaced helix at room temperature and 1 bar, when folding is initiated from the pressure-denatured state. Pressure unfolding experiments reveal a highly reversible denatured ensemble with helix-coil composition (Figure 4.4 and Appendix C, Figure C.6), and ultrafast P-jump refolding from this ensemble detects a 1.4-ms phase in addition to a microsecond phase (Figure 4.3). All-atom MD simulations show that after a P-jump, refolding trajectories can get stuck in structures with helix misplaced into loops for >8 μs. MD results for CHARMM27 and CHARMM22* are consistent with some fraction of the λ^*YA population being trapped in a state containing nonnative helix within $\leq$3 μs, from which it recovers in $\approx$1.4 ms according to experiment (possibly slowed further by transient aggregation). This refolding pathway is very different from refolding out of the thermally denatured state; in that state, the protein folds in <100 μs [242] and, if not, it either aggregates irreversibly or is trapped in a state with extended (β-like) structure, based on simulation [256] and thermal titrations [246] (Appendix C, Figure C.3).

Even for a helix bundle such as λ_{6-85}, there can be too much of a good thing. When nonnative helix encroaches on turns, it renders them unable to align the secondary structure elements so that they can assemble into the proper native tertiary structure. The "knotted" state in Figure 4.5 is a good example. Figure 4.6 highlights in red the turn/coil regions of λ^*YA that are particularly prone to forming nonnative helix. In both simulations, L31, S32, M42, G43, A56, L57, and N58 exhibit $\geq$80% average helical populations. According to the scale of Pace and Scholtz [270], the helical propensity for these residues is A>L$\approx$M>S>N>G. Thus, we propose that helix-breaking mutations L31G, M42G, A56G, and L57G in particular could reduce the transient trapping of λ-repressor fragment when it refolds from the P-denatured state.

Pressure denaturation of fast folders could provide a rich test bed for calibrating force fields against experimental data, by looking at transient nonnative structure in addition to native structure. The advantage of small fast folders in this regard is that such nonnative structures can form and dissolve on time scales accessible to full atom simulation, enabling comparisons of

mechanism beyond rate coefficients, stability, or native structure.

CHAPTER 5

MULTIPROBE MAPPING REVEALS THE FAST FOLDING MECHANISM OF λ-REPRESSOR FRAGMENT 6–85

5.1 Introduction

Unlike reactions of small organic molecules, it is difficult to accurately describe the process of protein folding by any single reaction coordinate. Nevertheless, comparisons of fast protein folding experiments and full-atom molecular dynamics simulations often rely on a single reaction coordinate that can be measured experimentally. Commonly used reaction coordinates include the radius of gyration, tryptophan fluorescence/lifetime, shift of the amide I′ band, change in FRET efficiency, *etc*. A single reaction coordinate would be suitable to describe an ideal case of two-state protein folding where the free energy of a protein can be accurately represented by two wells separated by a barrier significantly larger than the thermal energy. Although fast folding proteins are often assumed to fold in a two-state manner, recent experiments have shown that this assumption can be violated by a significant degree if on-pathway or off-pathway traps are present. Molecular dynamics simulations have also been able to resolve structural heterogeneity in the unfolded and trapped ensembles of fast folders and to reconstruct free energy landscapes that reveal complexity far beyond the assumed two-state scenario. Therefore it is important to examine fast protein folding using multiple probes and various perturbation methods in order to gain a more complete picture of the folding process and facilitate more precise comparison between simulations and experiments.

Another outstanding issue in the protein folding field is whether various perturbation methods alter the mechanism of protein folding by inducing different unfolded states. What the structure of the unfolded states is, is a complicated question because of the lack of structural analysis techniques to resolve it and therefore the inability to tweak the force fields to correctly

predict denatured conformations in full atomistic detail. Understanding the energetics of the unfolded state is important since misfolding happens when the native state first unfolds and then fails to revisit the native conformation. A single reaction coordinate is insufficient to understand the relationship between the folding mechanism and the perturbation variable because a change in folding rate might originate from the change in the folding barrier hight but not necessarily due to the change in the sequence of events that occur during the folding process. Moreover, care has to be taken when protein folding experiments with different perturbation variables are compared. For example, the final state of the perturbation has to be identical and the solvent conditions have to be the same.

Here, we attempted to generate a series of mutants of a fast folding five-helix bundle lambda repressor fragment 6-85 that would allow probing interactions between its helices during folding. We used tryptophan fluorescence quenched by tyrosine as a probe. In these experiments fluorescence lifetime of tryptophan depends on its distance from a tyrosine residue, which is intentionally placed in close proximity in the native state. Thus, in the native state, the fluorescence lifetime of tryptophan is short due to quenching by tyrosine and in the unfolded state the lifetime on tryptophan becomes longer. The mutants were designed empirically and their stability was measured against temperature, pressure, and guanidine hydrochloride, using fluorescence and circular dichroism spectroscopy as detection methods. Kinetics was then measured using temperature and pressure as perturbation variables. This study provides insights into the folding mechanism of lambda repressor, which provides a platform for further confirmation by molecular dynamics simulations. We also investigated the dependence of the folding mechanism of lambda repressor on the perturbation used to synchronize the ensemble and discovered that the folding mechanism is robust with respect to the perturbation method. These results may be useful in further developing a unified picture of protein folding in the pressure-temperature plane.

5.2 Materials and methods

5.2.1 Sample preparation

λ mutants were expressed as described previously.[225] The measurements were done in 50 mM K_3PO_4 buffer at pH = 7.1. Guanidine hydrochloride (GuHCl) was obtained from Sigma-Aldrich Inc. (St. Louis, MO).

5.2.2 Equilibrium folding experiments

CD temperature melts Circular dichroism (CD) thermal melts (Figures B.1–B.6, Table B.1) were done on a Jasco-715 spectropolarimeter (Jasco Inc., Easton, MD). The temperature was incremented by 3 °C from the initial value of 20 °C until the final temperature of 95 °C was reached resulting in 26 CD spectra. Temperature was controlled using a Peltier element. Each CD spectrum in figures is an average of 5 spectra at 100 nm/min scan speed. A quartz cuvette (Starna Cells Inc., Atascadero, CA) with 1 cm path length was used. All proteins were used at 2.5 μM concentration.

Fluorescence temperature melts Fluorescence spectra (Figure 5.2b) were measured using a fluorescence spectrophotometer Cary Eclipse (Varian Inc., Palo Alto, CA). The PMT voltage was usually 600-750 V, excitation and emission slit widths were 5 nm each, excitation wavelength was 280 nm, interval between points was 1 nm, and the scan rate was 600 nm/min. Sample concentrations were 10 μM for all proteins.

Fluorescence thermodynamics under pressure Fluorescence spectra were measured using a Cary Eclipse fluorescence spectrophotometer (Varian). Excitation and emission slit widths were 5 nm each, excitation wavelength was 280 nm, and the scan rate was 120 nm/min. Sample concentration was 200 μM. The sample was pressurized with a high-pressure cell (ISS). We used a rectangular quartz cuvette with a path length of 4 mm. Spectrophotometric grade ethyl alcohol (95.0%, A.C.S. reagent; Acros Organics) was used as pressurization fluid. The mean wavelength in Figure 5.5 is the weighted average $\langle I \rangle = \frac{\int d\lambda\ \lambda\ I(\lambda)}{\int d\lambda\ I(\lambda)}$ of the fluorescence intensity.[236].

5.2.3 Kinetics

T-jumps Laser temperature jump was done using a Surelite Q-switched Nd:YAG laser (Continuum Inc., Santa Clara, CA) Raman-shifted to 1.9 μm by passing the beam through a 1 m long tube with hydrogen gas pressurized to 300 psi. (Figures 3.1 and B.7) The beam was then passed through a 50% beam splitter so that the sample could be excited from two sides, which provided for more uniform heating of the sample during the temperature jump. The jump was usually 5-7 °C. The exact size of the jump was calibrated by comparing the fluorescence decays of tryptophan (300 μM solution) after the jump with the corresponding decays without the jump at the equilibrium temperature several degrees higher. Equilibrium temperature was set using an automated temperature controller (Lake Shore 330, Lake Shore Cryotronics Inc., Westerville, OH). The sample cell was made of fused silica tubing (3530S-100, VitroCom, Mountain Lakes, NJ) welded on one side. The excitation path length was 0.3 mm. The sample was excited with a Ti:sapphire laser (KMLabs Inc., Boulder, CO), which pulsed at 80 MHz frequency. The laser wavelength was 860 nm, which was frequency tripled with a third harmonic generator (CSK Optronics Inc., Torrance, CA) to 287 nm. Tryptophan fluorescence was then guided with an optical cable, passed through a B370 band-pass filter (Hoya Corp., Santa Clara, CA) and collected by a photomultiplier (R7400U-03, Hamamatsu Corp., Bridgewater, NJ). The signal was then recorded and digitized with an oscilloscope (DPO7254, Tektronix Inc., Beaverton, OR) with 2.5 GHz bandwidth. The length of the time traces was 5 ms and each trace contained 50 000 000 data points. Temperature jump was set to occur 153.75 μs after the oscilloscope was triggered to start data collection. The sampling frequency was 10 GS/s, which yielded a data point every 100 ps. The signal was usually 100-250 mV. The data were analyzed with software written in LabView (National Instruments Inc., Austin, TX) and Matlab (Mathworks Inc., Natick, MA). Sample concentration was 150 μM for all proteins, as measured by absorption at 280 nm.

P-jumps P-jumps with a home-built apparatus are described in detail elsewhere [235]. Briefly, 300 μM protein solution was pipetted into an $\sim$8-μL hemispherical dimple machined in-house into an optical grade 3/8-inch sapphire cube (Esco Products). The dimple was then covered with a double-layer

of mylar-coated aluminum foil to prevent mixing of the sample with ethanol, which served as the pressurization fluid. The cube was inserted into a pressurization clamp as described elsewhere [235], and the sample was pressurized to 1.2 kbar using a hydrostatic pump (High Pressure Equipment Company). We used 0.007-inch-thick stainless-steel burst membranes and 95 V (∼10-kA current) to burst the membrane. The sample was excited by a Ti:sapphire laser (KMLabs). The laser emission at 840–860 nm was frequency-tripled to 285 ± 3 nm. The laser was adjusted to a pulse rate of 80 MHz. Tryptophan fluorescence decay was collected by means of an optical waveguide (Oriel) every 12.5 ns and passed through a B370 band-pass filter (Hoya) onto a photomultiplier (R7400U-03; Hamamatsu Corp.). The fluorescence decays were recorded by a DPO7254 digitizer (Tektronix) with 2.5-GHz bandwidth, locked to the 80-MHz laser cavity to avoid aliasing. Each kinetic trace contained 400,000 fluorescence decays. Each fluorescence decay was sampled at 10 GHz (125 points per decay). The amplitude of the fluorescence signal was 100–250 mV. The data were analyzed using custom code written in LabView (National Instruments), MATLAB (MathWorks), and IGOR Pro (Wavemetrics). One hundred raw fluorescence decays were binned into average decays every 1.25 microseconds, reducing the time resolution to a value comparable to the dead time of the P-jump. To assign a single lifetime parameter to each fluorescence decay, we fitted them to the linear combination of a prejump and postjump fluorescence decay. We averaged the first 100 decays of the time trace to get a representative decay f_1 before the P-jump and the last 100 decays at 5 ms to get a representative decay f_2 after the P-jump. Decays between these two P-jumps were fitted by $f = a_1 f_1 + a_2 f_2$, and the fraction contributed by the second decay was calculated as $\chi = a_2/(a_1 + a_2)$. Thus, a decay with the same lifetime as before the jump yields $\chi = 0$, and a decay with the same lifetime as at 5 ms yields $\chi = 1$. We fit the resulting time trajectory to a double-exponential function $A_f e^{-t/\tau_f} + A_s e^{-t/\tau_s}$.

5.3 Results

5.3.1 Folding mechanism of lambda repressor fragment 6–85

We attempted to express four new lambda repressor mutants that would each have a pair of helices containing a tryptophan in one helix and a tyrosine in the other. Fluorescence of tryptophan in sensitive to its environment. In the folded state the tryptophan–tyrosine interaction would quench the fluorescence and shorten the fluorescence lifetime. In the unfolded state the fluorescence would be restored.

We named the mutants that we designed based on the following convention: λ_{WY} is a mutant of lambda with a tryptophan positioned in helix number W and a tyrosine positioned in helix number Y. We started with a Q33Y, Y22W, G46A, G48A mutant of the original lambda repressor fragment 6–85 (Figure 5.1a). This mutant has a tryptophan in helix 1 (position 22) and a tyrosine in helix 2 (position 33). We thus call it λ_{12}. The other three mutants were named accordingly (Figure 5.1b). We then calculated 100 ns of full-atom molecular dynamics optimization for each mutant and determined the distance between the tryptophan and the tyrosine (Figure 5.1c).

Two of the designed mutants, λ_{13} and λ_{12}, expressed very well (Figure 5.2d). λ_{42} exhibited moderate expression. In an effort to make a more stable mutant with helices 2 and 4 connected, we attempted to make a mutant λ_{42b} with a A62W mutation, instead of the L69W mutation in the original λ_{42} variant. Alanine is one turn of helix 4 closer to the N-terminus, while also facing towards the hydrophobic core. However, λ_{42b} showed even worse expression level than λ_{42}. λ_{41} also expressed poorly.

We proceeded with the experiments on the other three mutants: λ_{13}, λ_{32} and λ_{42}. We measured the temperature melting curves of these proteins as a function of temperature using fluorescence spectroscopy (Figures 5.2a and 5.2c) and circular dichroism spectroscopy (Figure 5.2b). The value of the mean wavelength is indicative of the solvent-exposure of the tryptophan. Our original protein, λ_{12}, showed the largest shift in mean wavelength upon unfolding. Interestingly, fluorescence spectrum of λ_{13}, which showed the largest CD intensity, was redshifted by $\sim$7 nm compared to λ_{12}. CD spectrum of λ_{42} did not exhibit cooperative behavior, indicating the deviation of lambda repressor from the two-state behavior based on the sensitivity of the signal

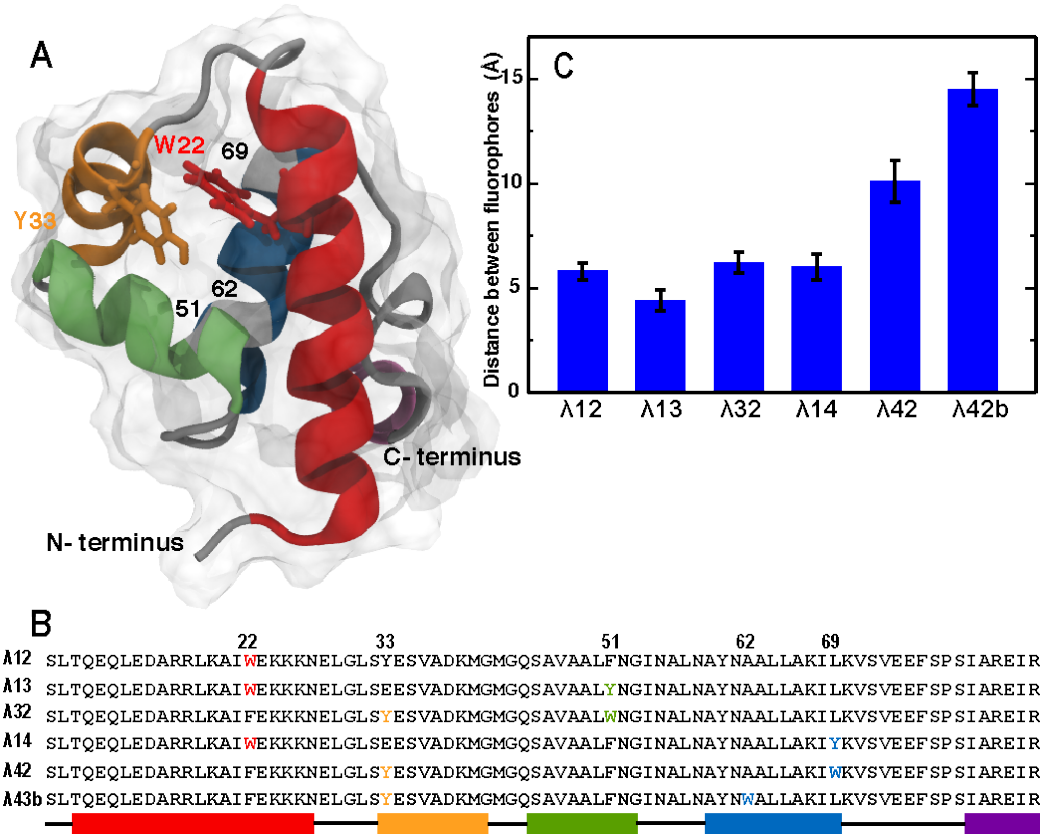


Figure 5.1: New lambda repressor mutants. (a) Crystal structure of lambda repressor mutant Q33Y, Y22W, G46A, G48A (PDB ID: 3KZ3). Helices 1–5 of the bundle are highlighted in red, yellow, green, blue and purple, respectively. W22 and Y33 are shown as stick models in red and yellow, respectively. The other mutations (F51, A62, and L69) are shown in gray. The Van der Waals surface of the protein is shown in transparent gray. (b) The sequences of the designed mutants. The mutants were named as λ_{WY}, where W is the helix in which tryptophan is located and Y is the helix in which tyrosine is located. λ_{42} is the L69W mutant, λ_{42b} is the A62W mutant. Alpha-helical regions are shown at the bottom. (c) Distances between W and Y for each mutant obtained from 100 ns molecular dynamics relaxations. Error bars represent standard deviations.

on the detection method. All proteins except λ_{42} exhibited an increase in fluorescence intensity when the temperature was increased, suggesting the anticipated proximity of tryptophan and tyrosine in the native state. For the reasons of low expression, lack of cooperatively as probed by CD and lack of tryptophan–tyrosine interaction as probed by fluorescence intensity, we abandoned the investigation of λ_{42} and proceeded with further analysis of the remaining two mutants, λ_{13} and λ_{32}.

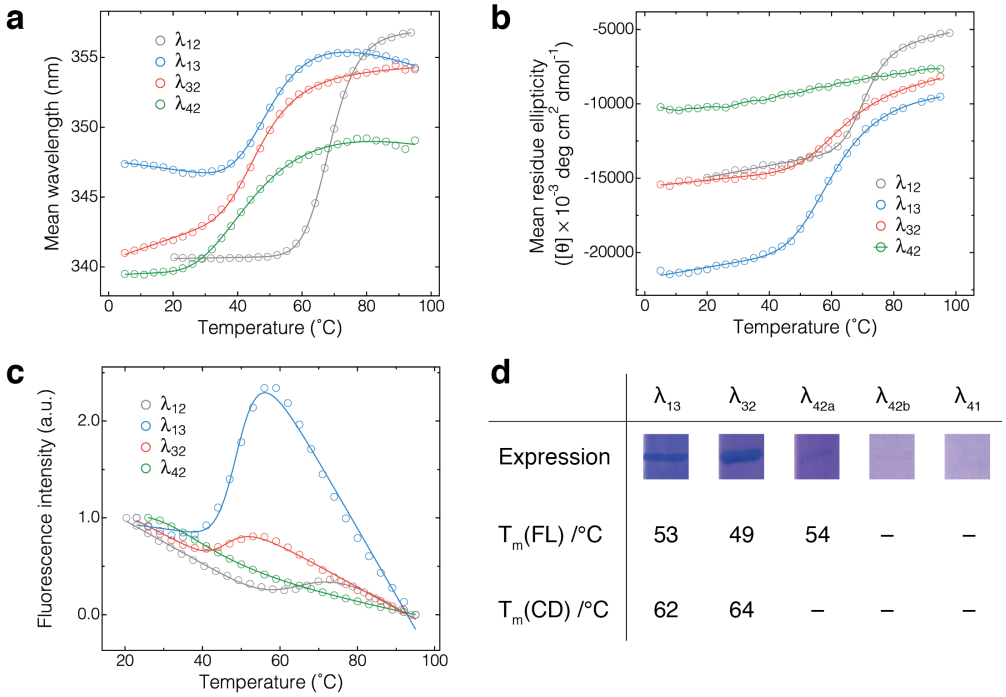


Figure 5.2: Stability of the nee lambda mutants. (a) Spectral mean of fluorescence spectra of the new lambda mutants as a function of temperature. λ_{12} is shown in gray for comparison. Solid lines are two-state thermodynamic fits. (b) Mean residue ellipticity at 222 nm as a function of temperature for the brew mutants. λ_{12} is shown in gray for comparison. Solid lines are two-state thermodynamic fits. λ_{42} did not show a cooperative transition and was not fitted. (c) Integrated fluorescence intensity normalized to begin at 1 and end at 0 for the new lambda repressor mutants. λ_{12} is shown in gray for comparison. Solid lines are two-state thermodynamic fits. Note that an increase in fluorescence intensity upon denaturation is observed for all mutants except λ_{42}, indicated a lack of W–Y interaction in the folded state. (d) Experssion efficiency and melting temperatures of the new lambda mutants. Acrylamide gels indicate good expression of λ_{13} and λ_{32}, satisfactory expression of λ_{42} and poor expression of λ_{42b} and λ_{41}.

We measured the kinetic response of λ_{13} and λ_{32} to a 9 °C upward temperature jump. In figure 5.3 we contrast the results we obtained with previously published data for λ_{12} and a truncated version of the protein, λ_{blue1}, which only has helices 1 and 4. We use $\lambda_{blue1}(14)$ as a proxy for λ_{41} due to the poor expression observed for λ_{41}.

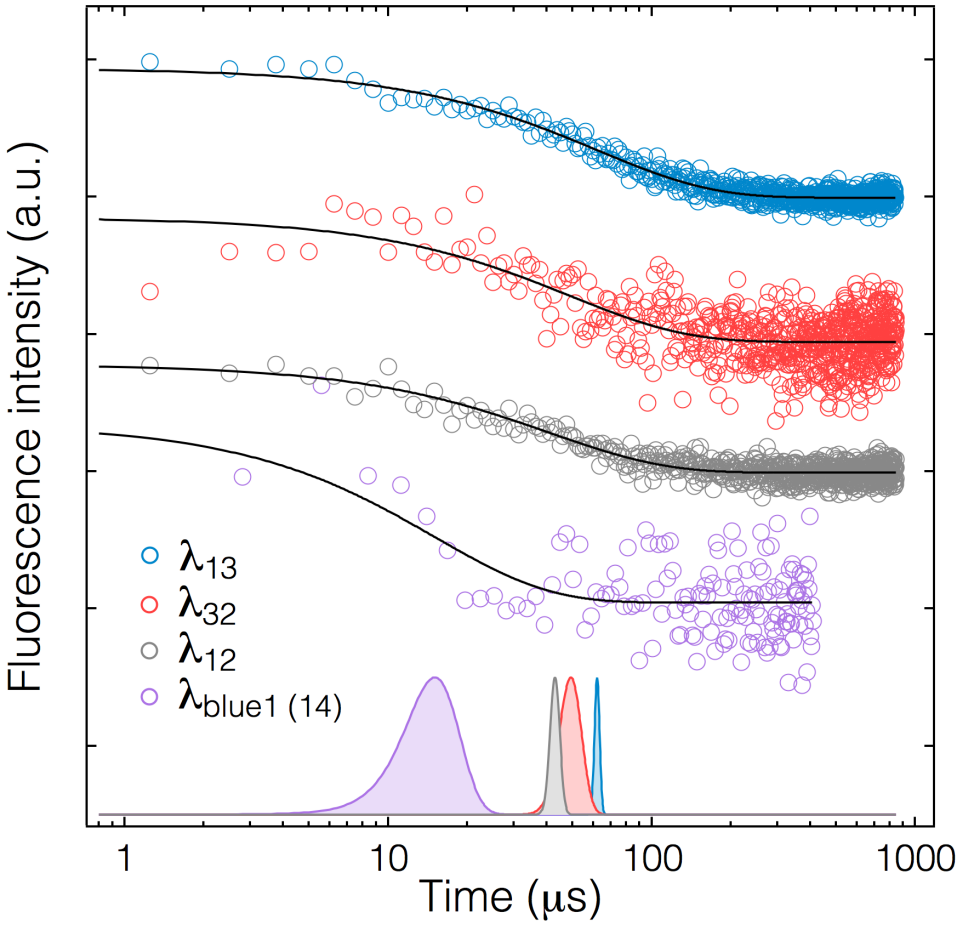


Figure 5.3: Kinetics of lambda repressor. Fluorescence lifetime normalized from 1 to 0 in arbitrary units and plotted against the logarithmic time axis. $\lambda_{blue1}(14)$ is from Prigozhin et al. (2011). λ_{12} is from Prigozhin et al. (2011). The solid black lines are single-exponential fits of the data: $I(t) = e^{-t/\tau}$. Gaussian profiles at the bottom correspond to the four decays above. The profiles are centered at the fitted τ_{obs} and have the width equal to twice the standard deviation.

5.3.2 The folding mechanism of lambda repressor fragment 6–85 is robust with respect to the perturbation variable

We also determined the stability of new lambda repressor mutants that we designed towards a chemical denaturant, guanidine hydrochloride (Figure 5.4). Both λ_{13} and λ_{32} showed very similar midpoint concentrations as analyzed by CD and fluorescence.

We sought to compare temperature-jump and pressure-jump kinetics of lambda repressor. To this end, we designed our kinetic experiments such that the final state of both jumps would be identical: 23 °C, 1 bar, 1.2 M GuHCl. The results are shown in figure 5.5. We did thermodynamic temperature and pressure melts of lambda mutants in 1.2 M GuHCl (Figures 5.5a and 5.5c). We then did temperature-jumps and pressure-jumps on lambda repressor mutants in 1.2 M GuHCl (Figures 5.5b and 5.5d). Pressure-jump showed slower kinetics than that observed with the temperature jump but the order of kinetic lifetimes was the same irrespective of the perturbation variable.

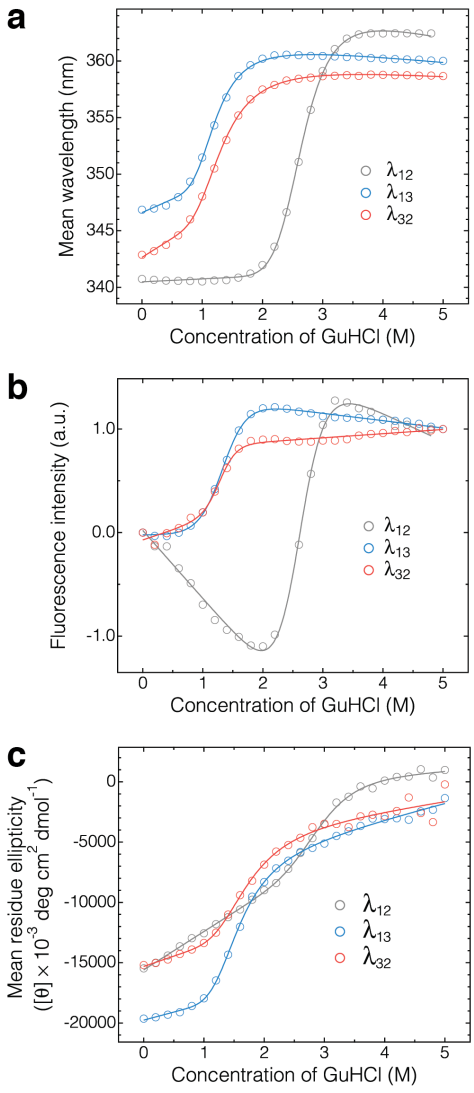


Figure 5.4: Stability of the new lambda repressor mutants as a function of guanidine concentration. (a) Spectral mean of fluorescence spectra of the new lambda mutants as a function of guanidine concentration. λ_{12} is shown in gray for comparison. Solid lines are two-state thermodynamic fits. (b) Integrated fluorescence intensity normalized to begin at 0 and end at 1 for the new lambda repressor mutants. λ_{12} is shown in gray for comparison. Solid lines are two-state thermodynamic fits. (c) Mean residue ellipticity at 222 nm as a function of guanidine concentration for the brew mutants. λ_{12} is shown in gray for comparison. Solid lines are two-state thermodynamic fits.

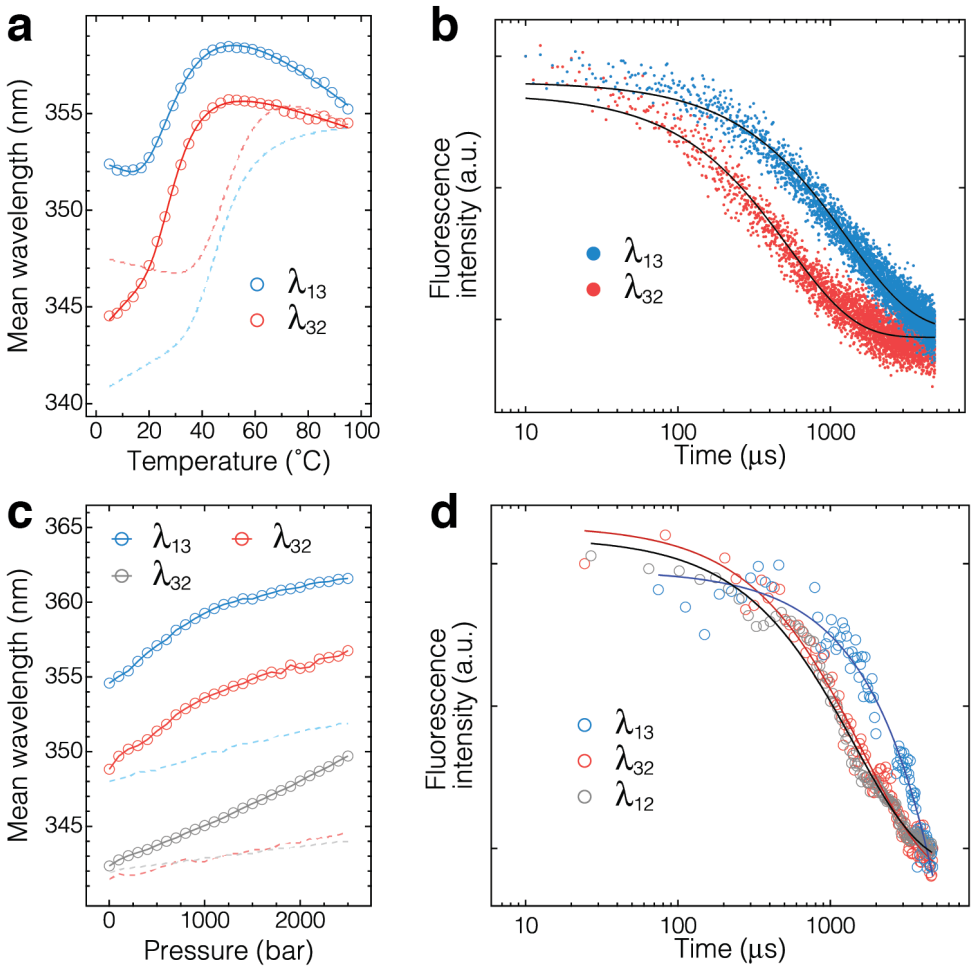


Figure 5.5: Robustness of the folding mechanism of λ_{6-85} with respect to the perturbation variable. (a) Spectral mean for fluorescence spectra of lambda measured in 1.2 M guanidine at 1 bar as a function of temperature (circles). Solid lines are two-state thermodynamics fits. Dashed lines are the melts at 0 M guanidine for each mutant shown for reference. (b) T-jump kinetics of the two mutants from 14 °C to 23 °C at 1.2 M guanidine at 1 bar. (c) Spectral mean for fluorescence spectra of lambda measured in 1.2 M guanidine at 23 °C as a function of pressure (circles). Dashed lines are the melts at 0 M guanidine for each mutant shown for reference. (d) P-jump kinetics of the two mutants from 1600 bar to 1 bar in 1.2 M guanidine at 23 °C.

5.4 Discussion and conclusion

We have investigated the details of the folding mechanism of lambda repressor fragment 6–85. As predicted previously by both experiments and simulations, the assertion that helices 1 and 4 come together first still holds true. According to our results, helix 2 then makes a contact with helix 1, followed by the docking of helix 3 to helix 2. Finally, helix 3 interacts with helix 1 to form the hydrophobic core of the protein. Our results further strengthen the idea that lambda repressor fragment 6–85 is not an idealistic two-state folder. Our findings provide a platform for further computational investigation of the mechanism of lambda repressor using millisecond timescale all-atom molecular dynamics simulations. Computational resources are now available not to assume that all mutants have the same equilibrium state and folding mechanism as the wild type, but rather to simulate each individual mutant and provide insight into the specific differences and similarities that the variants might share.

We have also discovered that the folding mechanism of lambda repressor fragment 6–85 is robust with respect to the perturbation variable. When the experimental conditions are carefully chosen so that the final state is the same in both experiments, the ordering of folding events does not change even though the individual observed rates are slower when pressure is used a perturbation variable as compared to temperature. These results provide a direct evidence to the fact that pressure-denatured state of lambda repressor 6–85 is different from the temperature-denatured state kinetically. However, the difference is not large enough to alter the folding mechanism *en route* to the native state. We hope that further computational investigations of the unfolded states under various denaturing condition will soon follow.

CHAPTER 6

KINETIC AND THERMODYNAMICS EXPLORATION OF THE PRESSURE-TEMPERATURE PHASE DIAGRAM OF PHOSPHOGLYCERATE KINASE

6.1 Introduction

Molecular systems are said to exhibit glassy behavior if they are kinetically trapped in local free energy minima. Such a rough energy landscape with free energy minima separated by kinetic barriers that are much larger than the available thermal energy is said to be frustrated. Similarly, the energy landscape theory of protein folding explains the folding process in the contest of an energy landscape, where high-energy high-entropy unfolded states interconvert with a low-energy low-entropy native state. This theory is closely related to the formalism used to explain the behavior of spin glasses. The difference comes from the fact that proteins fold on a minimally frustrated energy landscape, which means that the energy landscape is biased towards the native folded state and the large kinetic barriers common for glasses are smoothed. The glassiness of random heteropolimers made out of an alphabet of 20 units (20 amino acids) has been transformed into the propensity for efficient folding by evolution for most of the proteins in the cell. However, remnants of the inherent frustration can still be seen in the behavior of some proteins. These observations imply that modifications of the protein sequence through natural selection alone do not always prevent proteins from populating slow-interconverting trapped states. These misfolded states can be detrimental if their predisposition to form toxic oligomers and aggregates overpowers the proteostatisis network of the cell.

Is it then possible that the cell is using epigenetic mechanisms to get rid of the residual glassiness of some proteins? Here we show that macromolecular crowding can smooth out a region on the temperature-pressure free energy landscape of yeast phosphglycerate kinase (PGK), converting this protein

from a three-state to a two-state folder. Insights on the secondary and tertiary structures of the two states that collapse into one upon addition of a crowding agent, ficoll 70, were gained using FTIR spectroscopy in the amide I′ region and fluorescence spectroscopy using tryptophan as a probe. Our data indicates that the two glassy states are similar in terms of the secondary structure and solvent-exposure of the tryptophan that they exhibit. The two states persist at low temperatures (akin to the behavior of a glass below the glass transition temperature) but converge at a higher temperature, which is lower if ficoll 70 is present in solution.

It has been shown previously that macromolecular crowding can increase the enzymatic activity of PGK, which is much more important than just increasing the propensity of a protein to fold by destabilizing the high-entropy unfolded states. Our current results suggest that there may be yet another role that macromolecular crowding may play in the cell: we propose that macromolecular crowding can act to minimize the glassiness in biomolecular systems. It is not unreasonable to hypothesize that cells exploit this effect ubiquitously given the non-specific nature of macromolecular crowding.

6.2 Experimental methods

6.2.1 Sample preparation

PGK was expressed as described previously [251]. Ficoll PM 70 was obtained from Sigma-Aldrich and used without further purification.

6.2.2 Equilibrium experiments

Pressure denaturation monitored by fluorescence spectroscopy. Fluorescence spectra were measured using a Cary Eclipse fluorescence spectrophotometer (Varian). Excitation and emission slit widths were 5 nm each, excitation wavelength was 280 nm, and the scan rate was 120 nm/min. Sample concentration was 66 μM unless otherwise specified. The sample was pressurized with a high-pressure cell (ISS). We used a rectangular quartz cuvette with a path length of 4 mm. Spectrophotometric grade ethyl alcohol (95.0%,

A.C.S. reagent; Acros Organics) was used as pressurization fluid. Temperature denaturation monitored by fluorescence spectroscopy was done in a similar manner but using a temperature controller instead of a pressurization cell. For a temperature melt at 500 bar and for pressure melts at various temperatures, a water bath was used to control temperature and pressure simultaneously. The mean wavelength shown in figures is the weighted average $\langle I \rangle = \frac{\int d\lambda\ \lambda\ I(\lambda)}{\int d\lambda\ I(\lambda)}$ of the fluorescence spectra.[236].

Pressure denaturation monitored by infrared spectroscopy. Lyophilized protein was dissolved in 20 mM Na_3PO_4 buffer in D_2O (Sigma), and D_2O exchange was allowed to proceed for 2 h at room temperature and then overnight at 4 °C. The protein was then lyophilized again. For the measurement, the protein was dissolved in 20 mM Na_3PO_4 buffer in D_2O at pD = 7.0 to the final concentration of 1.7 mM. pH of deuterium (pD) was adjusted using DCl. Approximately 10 μL of the sample was then placed in a diamond anvil cell equipped with type IIa diamonds (High Pressure Diamond Optics). The sample was held in a stainless-steel gasket (overall diameter = 12.5 mm, center pinhole diameter = 0.45 mm, thickness = 0.050 mm). The gasket was secured in the diamond anvil cell using gum from Faber–Castell (127020). A small amount of $BaSO_4$ was placed in the chamber with the sample. IR spectra were collected using a Nicolet Magna IR 550 spectrometer equipped with a nitrogen-cooled mercury cadmium telluride (MCT) detector. The sample chamber within the instrument was purged with dry CO_2-free air. The IR beam was focused onto the pinhole of the diamond anvil cell. Spectra were collected from 400 to 4,000 cm^{-1} using 256 accumulations and a resolution of 1 cm^{-1}. Pressure was increased using the spring-loaded screw of the diamond anvil cell assembly. Changes in the pressure were quantified by monitoring the stretching vibration of the calibrant, $BaSO_4$. Its peak was at 983.6 cm^{-1} at 1 bar and shifted linearly with pressure toward larger wave numbers.

6.2.3 Pressure-jumps

Microsecond-to-millisecond-scale pressure jumps. P-jumps with a home-built apparatus are described in detail elsewhere [235]. Briefly, 300 μM protein solution was pipetted into an ∼8-μL hemispherical dimple machined in-house into an optical grade 3/8-inch sapphire cube (Esco Products). The dim-

ple was then covered with a double-layer of mylar-coated aluminum foil to prevent mixing of the sample with ethanol, which served as the pressurization fluid. The cube was inserted into a pressurization clamp as described elsewhere [235], and the sample was pressurized to 1.2 kbar using a hydrostatic pump (High Pressure Equipment Company). We used 0.007-inch-thick stainless-steel burst membranes and 95 V (∼10-kA current) to burst the membrane.

The sample was excited by a Ti:sapphire laser (KMLabs). The laser emission at 840–860 nm was frequency-tripled to 285 ± 3 nm. The laser was adjusted to a pulse rate of 80 MHz. Tryptophan fluorescence decay was collected by means of an optical waveguide (Oriel) every 12.5 ns and passed through a B370 band-pass filter (Hoya) onto a photomultiplier (R7400U-03; Hamamatsu Corp.). The fluorescence decays were recorded by a DPO7254 digitizer (Tektronix) with 2.5-GHz bandwidth, locked to the 80-MHz laser cavity to avoid aliasing. Each kinetic trace contained 400,000 fluorescence decays. Each fluorescence decay was sampled at 10 GHz (125 points per decay). The amplitude of the fluorescence signal was 100–250 mV.

The data were analyzed using custom code written in LabView (National Instruments), Matlab (MathWorks), and IGOR Pro (Wavemetrics). One hundred raw fluorescence decays were binned into average decays every 1.25 microseconds, reducing the time resolution to a value comparable to the dead time of the P-jump. To assign a single lifetime parameter to each fluorescence decay, we fitted them to the linear combination of a prejump and postjump fluorescence decay. We averaged the first 100 decays of the time trace to get a representative decay f_1 before the P-jump and the last 100 decays at 5 ms to get a representative decay f_2 after the P-jump. Decays between these two P-jumps were fitted by $f = a_1 f_1 + a_2 f_2$, and the fraction contributed by the second decay was calculated as $\chi = a_2/(a_1 + a_2)$. Thus, a decay with the same lifetime as before the jump yields $\chi = 0$, and a decay with the same lifetime as at 5 ms yields $\chi = 1$. We fit the resulting time trajectory to a double-exponential function $A_f e^{-t/\tau_f} + A_s e^{-t/\tau_s}$.

6.3 Results

6.3.1 Pressure-temperature phase diagram of phosphoglycerate kinase is three-state at low temperatures

We used a destabilized mutant of PGK, which included mutations Y122W, W308F and W333F of the wild type protein. We explored the pressure-temperature free energy landscape of phosphoglycerate kinase (Figure 6.1a). Pressure denaturation of PGK was monitored by fluorescence at six different temperatures: 9, 15, 23, 30, 36, and 44 °C. The sample was pressurized using a commercial pressure cell and and hydrostatic pressure pump. Fluorescence spectra were recorded from 1 bar to 2500 bar in 100 bar increments (Figure 6.1b). Fluorescence probe that was monitored in this study is Y122W. PGK has a very blue-shifted fluorescence in its folded state with a peak wavelength of ∼306 nm. Upon pressure denaturation fluorescence shifts dramatically towards the redder region of the spectrum with the spectrum at 2500 bar peaked at 344 nm. The bandwidth of the spectrum also increases drastically from 47 nm at 1 bar to 85 nm at 2500 bar.

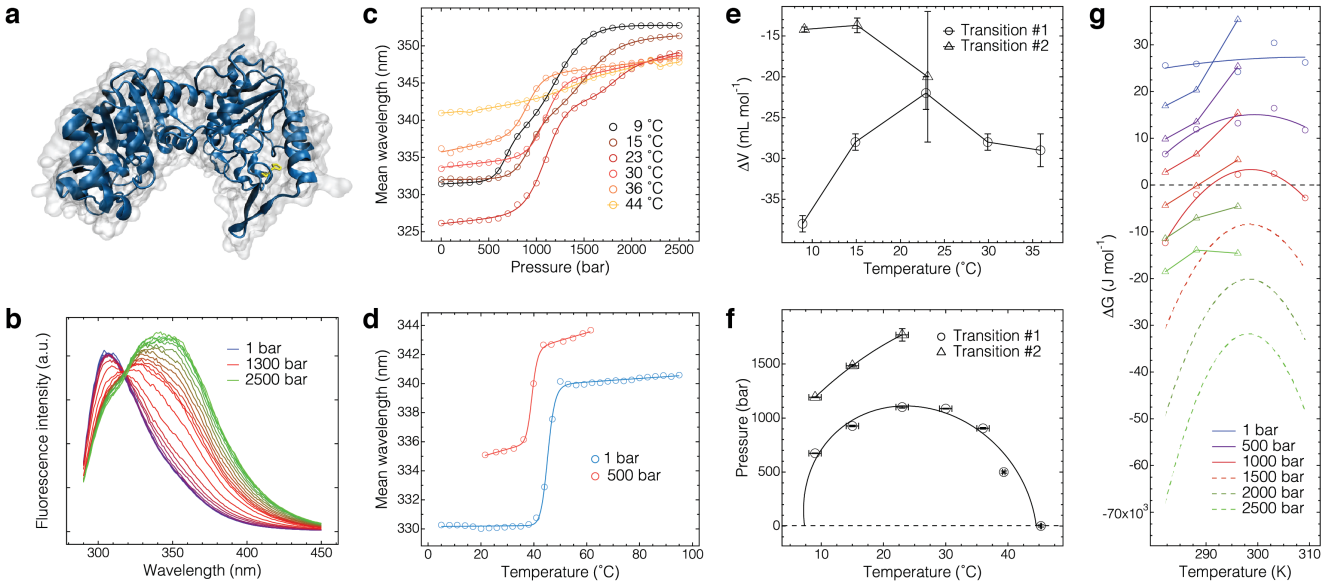


Figure 6.1: Behavior of phosphoglycerate kinase in the pressure-temperature plane. (a) Crystal structure of wild type phosphoglycerate kinase (PDB ID: 1QPG). The backbone of the protein is shown in blue, the Van der Waals surface is shown in transparent gray and Y122 (W122 in the destabilized variant) of domain 1 is shown in yellow. (b) Fluorescence spectra of PGK taken in increments of 100 bar at 23 °C from 1 bar to 2500 bar. (c) Spectral mean of the spectra (similar to the ones plotted in (b) at various temperatures. The solid lines represent two- or three-state thermodynamic fits. Note that three thermodynamic states are observed at 9, 15, and 23 °C. Profiles at 30 and 36 °C show two-state behavior. The protein is denatured at 44 °C. (d) Spectral mean of temperature melts of PGK at 1 and 500 bar. The solid lines represent two-state thermodynamic fits. (e) Temperature dependence of the negative change in volume between the native state and the intermediate state (circles, transition 1), and between the intermediate state and the denatured state (triangles, transition 2). (f) Pressure temperature-phase diagram of PGK. Circles represent midpoint pressures (at 9, 15, 23, 30, 36 °C) or temperatures (at 1 and 500 bar) for transition 1. Triangles represent midpoint pressures for transition 2. Solid elliptical curve going through the circles is a fit representing the $\Delta G = 0$ curve. A curve going through the triangles is a guide for the eye. (g) Free energy change profiles for transition 1 (circles) and transition 2 (triangles) as a function of temperature at various pressures. Solid lines going through the circles are the van't Hoff equation fits. Dashed lines are the extrapolated van't Hoff fits with a constant allowing for the negative apex of the parabola.

Fluorescence spectra were analyzed to extract the spectral mean (Figure 6.1c, see Methods), which is indicative of the spectral shift of the fluorescence spectra as a function of pressure. The resulting curves were fitted using three-state and two-state thermodynamic models (solid lines in Figure 6.1c). Two transitions were observed at 9, 15, and 23 °C. The transition at lower temperature is called transition 1 and the transition at higher pressure is called transition 2 throughout this chapter. As the temperature increases, transition 1 decreases in amplitude until it eventually disappears completely and only one transition is observed at 30 and 36 °C. The protein is unfolded at 44 °C and no transition is observed. In order to determine the transition points in the pressure-temperature plane at high temperatures (where pressure melts did not result in a cooperative transition because the protein was largely unfolded), temperature melts at 1 and 500 bar were done (Figure 6.1d). These melts were both fitted to a two-state model because only one transition was observed at high temperature.

Pressure and temperature melts shown in figure 6.1c were fitted to thermodynamic models where the following assumption was made to determine the dependence of free energy on pressure: $\Delta G = \Delta G_o + P\frac{\partial \Delta G}{\partial P} = \Delta G_o + P\Delta V$. We shifted the x-axis to force $\Delta G = 0$ at the melting pressure, P_m, which yielded: $\Delta G = (P - P_m)\frac{\partial \Delta G}{\partial P} = (P - P_m)\Delta V$. This general equation was used to fit both transition 1 and transition 2 as ΔG_{NI} and ΔG_{IU}, respectively, where ΔG_{NI} is the difference in free energy between the native state and the intermediate state, ΔG_{IU} is the difference in free energy between the intermediate state and the unfolded state. This Taylor expansion of the free energy was used to determine the depend of the change in volume, ΔV (Figure 6.1e), and the change in free energy, ΔG (Figure 6.1g), on temperature for the two transitions. It also allowed us to generate a pressure-temperature phase diagram of PGK (Figure 6.1f), where each data point is a midpoint pressure or temperature, *i.e.* $\Delta G = 0$, for the transition in question. Transition 1 was then fitted to an elliptical curve, which is common for the pressure-temperature phase diagrams of two-state folders:

$$\Delta G_u(P,T) = -\frac{\Delta\beta}{2}(P-P_m)^2 - \frac{\Delta C_p}{2T_o}(T-T_m)^2 + \Delta\alpha(P-P_m)(T-T_m) + \Delta V_o(P-P_m) - \Delta S_o(T-T_m) + \Delta G_o = 0 \quad (6.1)$$

Transition 2 was not observed past 23 °C within the range of pressures available in our experiment (1–2500 bar).

ΔG as a function of temperature at various pressures was plotted in figure 6.1g for both transitions. The data for transition 1 was fitted using the van't Hoff equation:

$$\Delta G_u(T) = \Delta H_u(T_m)\left(1-\frac{T}{T_m}\right) + \Delta C_{P,u}(T-T_m) + \Delta C_{P,u}T\ln\left(\frac{T_m}{T}\right) \quad (6.2)$$

The curves at pressures above 1000 bar were fitted with a modified version of equation 6.2, where a constant was added to allow for a negative apex of the parabola to highlight the change in curvature of the parabola as the pressure goes up. The result indicates that the depend on temperature becomes stronger as the pressure increases. Such a dependence is expected for the unfolding of the native state.

6.3.2 Effect of ficoll 70 on pressure stability of phosphoglycerate kinase

We analyzed the effect of ficoll 70 on the pressure stability of phosphoglycerate kinase. The spectral mean profiles at 23 °C at various ficoll concentrations are shown in figure 6.2a. The three-state scenario observed in the absence of ficoll at 23 °C (also shown in figure 6.1c and reflected in the pressure-temperature phase diagram of PGK in figure 6.1f) disappears upon the addition of ficoll even at 50 mg/mL and the two-state thermodynamic transition emerges. These transitions linearly shift towards larger pressure (Figure 6.2b) as more ficoll is added. They also become more cooperative as a function of ficoll concentration, which is expected based on the previous reports for ribonuclease A. The increase in protein stability (attributed to the destabilization of the high-entropy unfolded state) is also reflected in the

change of free energy as a function of ficoll at various pressures (Figure 6.2d).

6.3.3 Effect of ficoll 70 on the pressure-temperature phase diagram of phosphoglycerate kinase

The same procedures as described in section 5.3.1 were employed to obtain a pressure-temperature phase diagram of PGK in the presence of 100 mg/mL ficoll 70 (Figure 6.3). As expected, ficoll stabilized PGK not only at 23 °C as shown in section 5.3.2, but also at other temperatures. However, the surprising result lies in the shift from the three-state to the two-state behavior at 23 °C upon the addition of ficoll (Figures 6.3c and 6.3f).

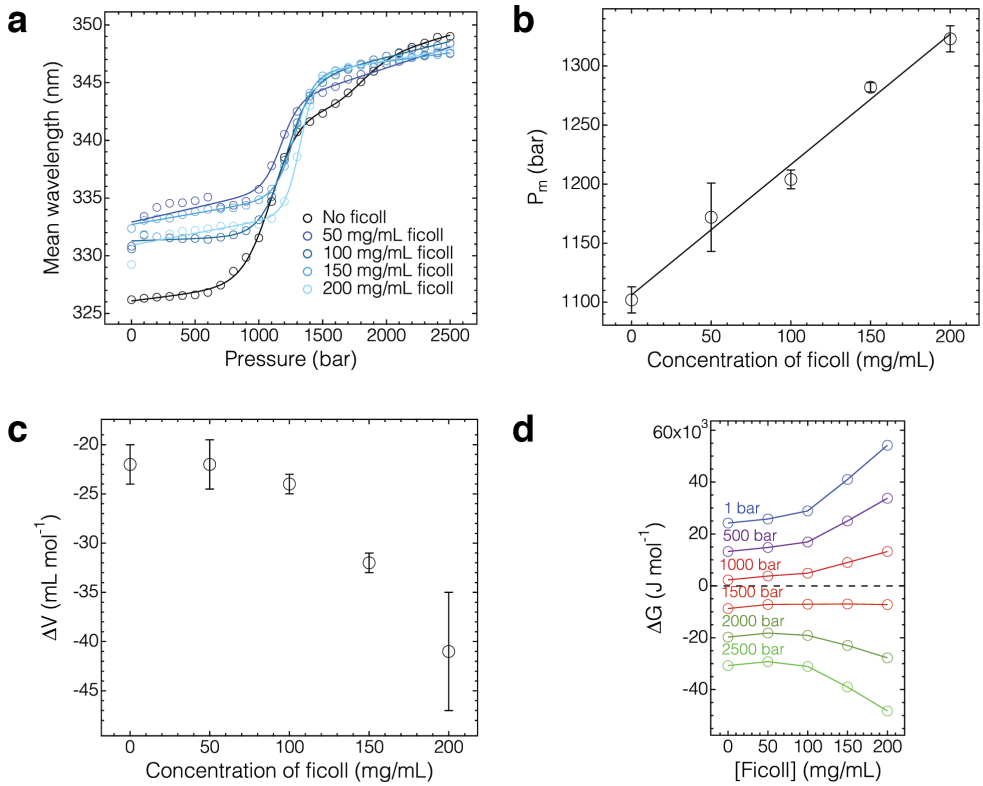


Figure 6.2: Pressure-induced denaturation of phosphoglycerate kinase at various ficoll 70 concentrations. (a) Spectral mean of the spectra (similar to the ones plotted in 6.1b and 6.3b) at various ficoll concentrations. T = 23 °C. The solid lines represent two- or three-state thermodynamic fits. Note that three thermodynamic states are observed only when no ficoll is present. Profiles at 50, 100, 150, and 200 mg/mL all show two-state behavior. (b) Midpoint pressures obtained from the fits to data shown in panel (a) plotted as a function of ficoll concentration. Only midpoint pressures of transition 1 are plotted. Transition 2 only appears when no ficoll is present. The solid line is a linear fit of the data. (c) Negative change in volume between the native state and the intermediate state (transition 1) as a function of ficoll concentration. (d) Free energy change profiles for transition 1 as a function of ficoll concentration at various pressures.

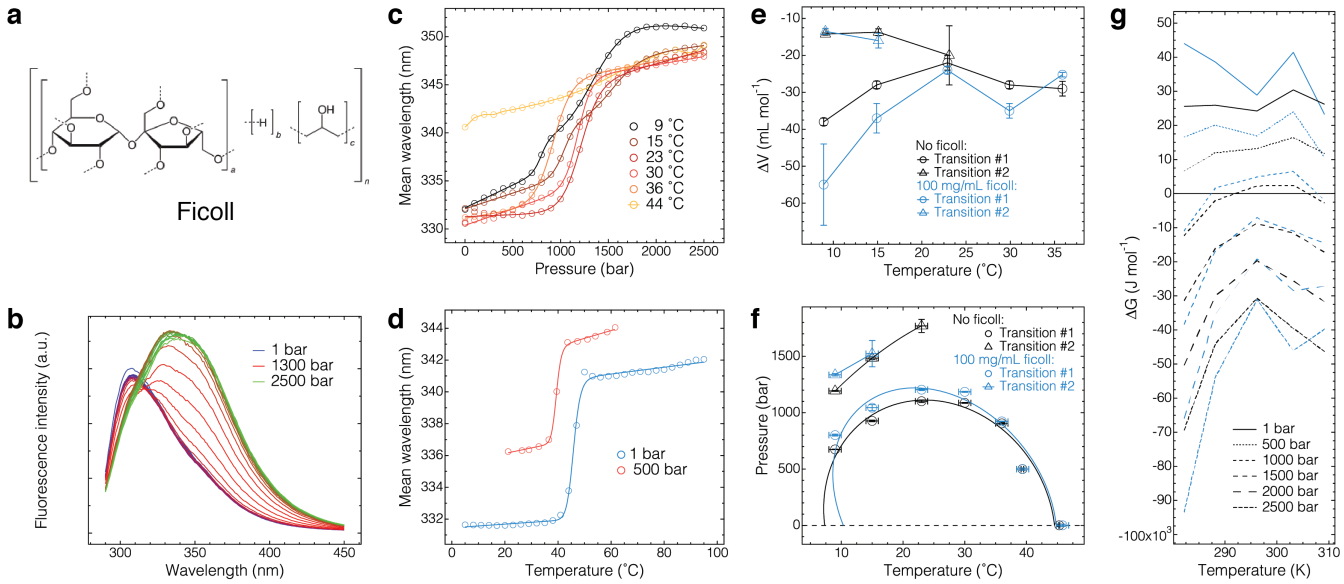


Figure 6.3: Behavior of phosphoglycerate kinase in the pressure-temperature plane in the presence of 100 mg/mL ficoll 70. (a) Ficoll 70 is a polymer with molecular weight $\approx$ 70 kDa. (b) Fluorescence spectra of PGK crowded by ficoll taken in increments of 100 bar at 23 °C from 1 bar to 2500 bar. (c) Spectral mean of the spectra (similar to the ones plotted in (b) at various temperatures ([ficoll] = 100 mg/mL). The solid lines represent two- or three-state thermodynamic fits. Note that three thermodynamic states are observed at 9, 15 °C, not at 23 °C as was the case without ficoll. Profiles at 23, 30 and 36 °C show two-state behavior. The protein is denatured at 44 °C. (d) Spectral mean of temperature melts of PGK at 1 and 500 bar ([ficoll] = 100 mg/mL). The solid lines represent two-state thermodynamic fits. (e) Temperature dependence of the negative change in volume between the native state and the intermediate state (circles, transition 1), and between the intermediate state and the denatured state (triangles, transition 2). Black – no ficoll, blue – 100 mg/mL ficoll. (f) Pressure temperature-phase diagram of PGK. Black – no ficoll, blue – 100 mg/mL ficoll. Circles represent midpoint pressures (at 9, 15, 23, 30, 36 °C) or temperatures (at 1 and 500 bar) for transition 1. Triangles represent midpoint pressures for transition 2. Note that transition 2 is not observed in the presence of ficoll at 23 °C. Solid elliptical curves going through the circles are the fits representing the $\Delta G = 0$ curves. Curves going through the triangles are guides for the eye. (g) Free energy change profiles for transition 1 as a function of temperature at various pressures with (blue) and without (black) ficoll.

6.3.4 Characterization of secondary structure of the unfolded states with and without ficoll

We used FTIR spectroscopy in the amide I′ region to gain insight into the secondary structures that are experienced by the non-native states observed at 23 °C under pressure (Figure 6.4). We normalized the spectra collected without ficoll in solution to go from 1 to 0 in the parameter $\chi(P)$ (see Data Analysis). Spectra taken in the presence of ficoll were analyzed using the basis functions obtained from the no-ficoll case, which leads to a baseline value of $\sim$ -0.6. Moreover, transitions observed with infrared spectroscopy are more two-state like both with and without ficoll. In both cases spectral broadening is observed at high pressure, which has previously been attributed to the denaturation of the protein and the increase in the β-content of the pressure-denatured conformation.

6.3.5 Pressure-jump kinetics of phosphoglycerate kinase

We did pressure-jumps on PGK ranging the time scales from microseconds to hours. We did not observe any fast (μs–ms) kinetic phases following a 1600 bar downward pressure jump as monitored by by fluorescence lifetime (Figure 6.5a). On the minute–hour timescale, a τ_{obs} $\approx$7 min kinetic phase is observed at 0, 50 and 100 mg/mL ficoll concentration. This fast phase persists but slows down as the ficoll concentration is raised from 100 to 200 mg/mL. In addition, a slow timescale appears at 100 mg/mL ficoll concentration, which is also retarded by the increasing ficoll content.

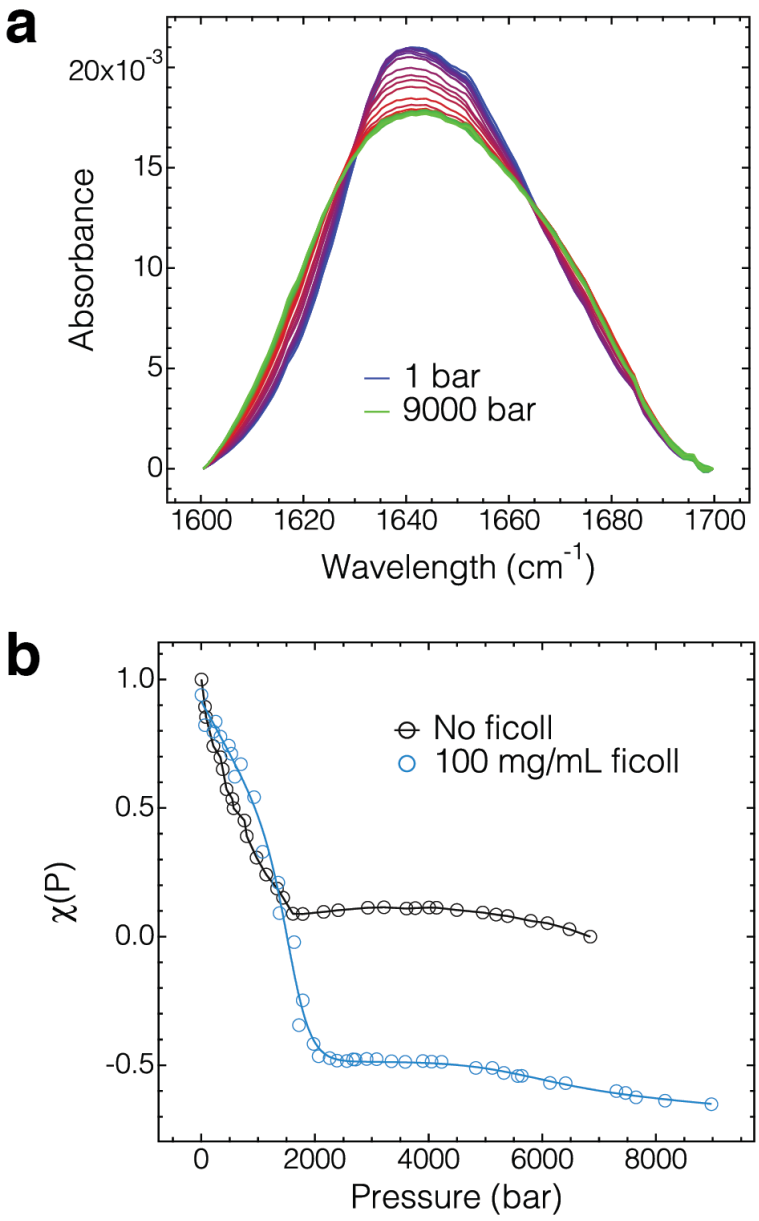


Figure 6.4: Pressure melts of PGK with and without ficoll probed by infrared spectroscopy in the amide I′ region. (a) Infrared spectra of PGK crowded by ficoll (100 mg/mL) at 23 °C from 1 bar to 9000 bar. (b) Infrared spectra of PGK under pressure without ficoll (black) were linearly normalized to go from 1 to 0 as described in the Data Analysis section. Spectra of PGK crowded by 100 mg/mL ficoll were then analyzed using the basis functions of spectra without ficoll to obtain data points in blue. The solid line going through the blue circles is a three-state thermodynamic fit.

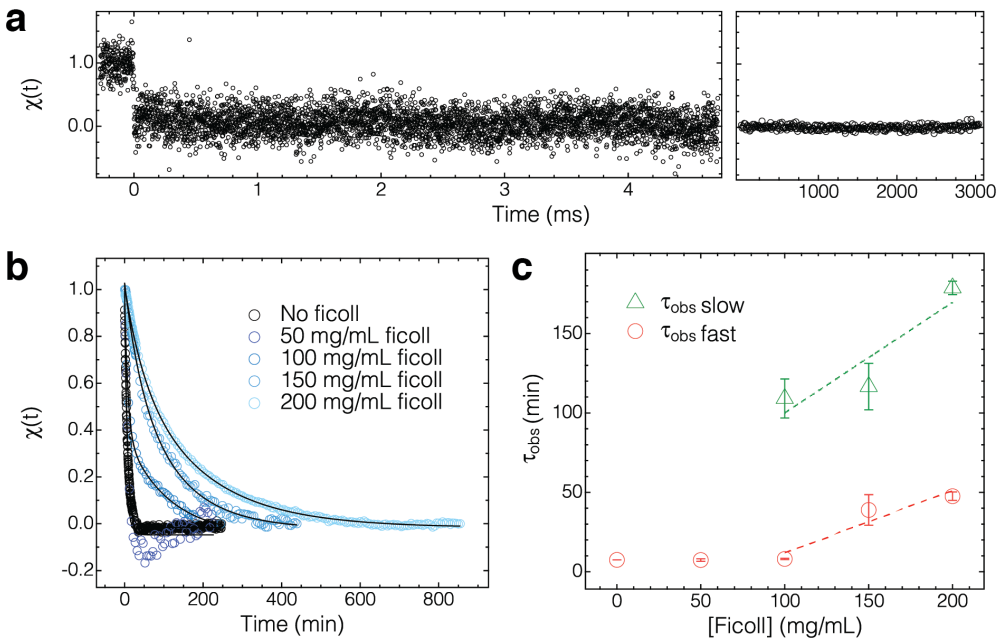


Figure 6.5: Pressure-jump kinetics of phosphoglycerate kinase with and without ficoll 70. (a) Pressure-jumps of PGK from 1600 bar to 1 bar. T = 23 °C. The panel on the left shows a microsecond-to-millisecond time scale experiment, 1 bin = 1.25 μs. The panel on the right shows a millisecond-to-second time scale experiment, 1 bin = 10 ms. The data are normalized from 1 to 0 based on the recorded fluorescence lifetime (see Data Analysis). (b) Pressure-jumps of PGK from 900 to 1000 bar at various ficoll 70 concentrations. The data points were generated using the spectral mean analysis and normalized from 1 to 0. T = 23 °C. Solid lines are single-exponential (for [ficoll] = 0 and 50 mg/mL) or double-exponential (for [ficoll] = 100, 150 and 200 mg/mL) fits. (c) Observed time constants, τ_{obs}, from the fits in (b). Dashed lines are guides for the eye.

6.4 Discussion and conclusion

6.4.1 Structures of the two non-native states

The lack of second transition in IR means that the two non-native states have a similar secondary structure, which leads us to believe that these are two partially unstructured states. The difference in $\chi(P)$ baseline of IR spectra with and without ficoll suggests that the unfolded states with and without ficoll are different and the addition of ficoll influences both of them. But the lack of the transition in fluorescence also indicates that it is probably one state in the presence of ficoll and two different states with similar secondary structure content in the absence of ficoll.

Currently there is some experimental evidence that the lower pressure unfolded state has some residual PP-II (beta-like) structure based on IR data, whereas the higher pressure denatured state may be more coil-like and exposes Trp 122 in the N domain. In addition, the lower P denatured state may become the dominant state at high T. We found previously that high-T high-denaturant unfolded states of proteins show evidence of PP-II structure, which actually has higher entropy than a coil because it allows the side chains more freedom to move around. Such transient PP-II strutcure could be an attractor pre-formed in the monomer, for forming beta-aggregates. Coarse-grained molecular dynamics simulations of PGK under pressure might be able to provide stronger structural constraints on the nature of the observed non-native states.

6.4.2 The role of macromolecular crowding to minimize the glassiness of bimolecular systems

Our results suggest that ficoll does not simply stabilize the native state or stabilize the non-native conformation that form when the native state is denatured under pressure. It creates a new non-native conformation, which is not the same as the conformations that are present in the absence of ficoll. The new conformation has a uniform signature as seen by IR and fluorescence at pressure where the native state is unfolded, suggesting that only one state is present and this one state is unique to ficoll-containing solution.

There is a difference between simply stabilizing one non-native state and

merging the two non-native states together. Stabilization of one non-native state over the other does not eliminate the possibility of the less stable state being populated at least for some of the time. In this case, the proteostasis network of the cell is put under the pressure of dealing with two different denatured states, one or both of which can be prone to oligomer formation and subsequent aggregation. Macromolecular crowding may help solving this problem but allowing only one denatured state to be present. By putting structural constraints on the unfolded states that can be present in a cell, macromolecular crowding may be alleviating the job of non-specific chaperones that protect the cell during stress response. It is easy to see how this effect may be ubiquitously employed by the cells given the promiscuous nature of macromolecular crowding.

CHAPTER 7

MECHANICAL MODELING AND COMPUTER SIMULATION OF PROTEIN FOLDING

7.1 Introduction

Simple models are used ubiquitously in science to describe the behavior of complex systems.[271–273] The Next Generation Science Standards[274] have emphasized "Developing and Using Models" as one of 8 core science and engineering practices that all students should learn. Key aspects of this practice include a) learning to represent and explain phenomena with multiple types of models and move flexibly between model types, b) being able to recognize and articulate limitations of models and describe how they might be improved to better fit available evidence, and c) being able to use models to investigate systems which are not readily available to the naked eye. Each of these skills is emphasized in the activity presented here.

The purpose of simple models is to reduce the dimensionality of a problem so that general trends can be deduced and the behavior of similar systems can be predicted.[275] A scientific model faces the issue of the balance between simplicity and accuracy. At one extreme, the model can account for nearly every detail and degree of freedom. While such a description will yield accurate results for specific systems, the large number of parameters can make the model inefficient and difficult to conceptualize. At the other extreme, the model is oversimplified and does not provide accurate results. There are several examples where such reductionist descriptions are very useful (*e. g.* ideal gas law) but many of these simplistic formalisms are too limited in their ability to explain a wide range of behaviors, mainly due to broad assumptions used to allow for an extreme reduction in the number of variables. Albert Einstein is often credited with the following quote: "Everything should be as simple as possible, but not simpler." Thus, the spectrum of scientific models

varies in its predictive power and utility as a function of coarse-graining, with the most functional and reliable ones being somewhere in between the two extreme scenarios as described in Figure 7.1.

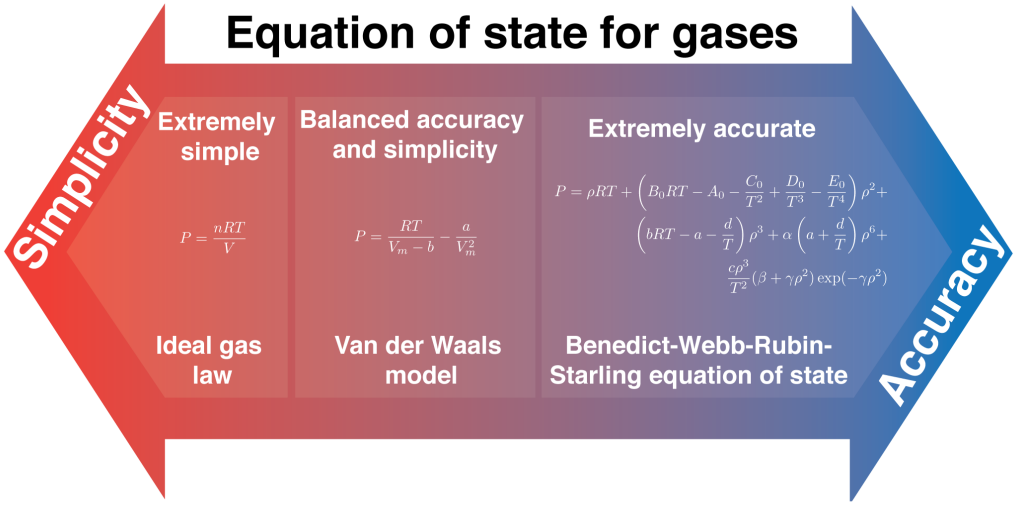


Figure 7.1: Equations of state for gases as a function of simplicity and accuracy of each model. The simplest model is the ideal gas law (on the left), while one of the more accurate ones is the Benedict-Webb-Rubin-Starling (BWRS) equation of state (on the right). The two extremes are balanced by introducing two (instead of twelve in the BWRS equation!) extra parameters into the ideal gas law, which results in a formula that we know as a Van der Waals model (in the middle).

In introductory chemistry, many of the models students are introduced to are simplistic and only provide accurate descriptions for a small subset of cases. These models reside at the "simple but inaccurate" end of the spectrum. While such models are useful for conceptual understanding of the phenomena that they describe, their use must be accompanied by a discussion of their limitations. Students often believe that models are reflections of reality, rather than having the more nuanced understanding that models are tools for explaining phenomena and have variable conceptual benefits and weaknesses. This view does not improve simply as a function of more educational experience and students need the opportunity to apply multiple models to the same system in order to develop a more sophisticated understanding of the use of scientific models.[276, 277]

We approached this problem by first introducing an asymptotic case of a very simple model and then systematically increasing its complexity to make

it more realistic. This interplay between accuracy and simplicity as well as strategies for balancing the two are at the core of this activity.

7.2 The model

The protein folding problem serves as a background concept for this activity. The protein folding problem constitutes a broad field of research in macromolecular chemistry, which contains all the aspects of thermodynamics and kinetics that a traditional "heat-engine-type" physical chemistry class would include. This area of biophysical chemistry is so rich that entire physical chemistry classes for biochemistry majors based solely on the protein folding literature were proposed.[278] Stand-alone lessons on protein folding have also been suggested to challenge the traditional use of predominantly inorganic chemistry material in general chemistry classes and to cater to the interests of students interested in pursuing health-related careers.[279–281]

The protein folding problem can be summarized in two questions: 1) How can the sequence of amino acids that comprises the protein uniquely define the native state of that protein, and 2) How can this unique native state be reached in a matter of microseconds when the number of possible conformations that a protein could randomly sample is so large? At the level of an introductory chapter in a biochemistry textbook, a protein is usually visualized as a rigid body made up of spiral ribbons (alpha-helices) and curved arrows (beta-sheets). This picture touches on question 1 of the protein folding problem and completely ignores question 2, falling at the simple end of the modeling spectrum. Increasingly accurate, yet less simple, models are required to better understand the protein folding problem.

Chemists often use Gibbs free energy surfaces to describe the stability of thermodynamic systems and the kinetics of the processes that describe a change between chemical states. A free energy surface for the simple rigid picture of a protein described above would look like a very narrow potential well, where only the native (folded) state rigidly defined by crystallographic assignment has any significant probability of being populated (Figure 7.2A). A slightly more sophisticated perspective can be achieved by allowing the protein molecule to wiggle around the crystallographic state in compliance with the thermal fluctuations of the medium. This flexibility would result

in the broadening of the free energy well (Figure 7.2B). Furthermore, if the solvent conditions do not strongly favor the native state, a protein could fold and unfold, exploring two free energy wells connected by a barrier (Figure 7.2C). Only at this point of model sophistication does the free energy surface resemble those taught in introductory thermodynamics. Ultimately, the native environment for proteins is not a buffer solution with carefully adjusted composition; it is the heterogeneous milieu of the living cell with all of its non-specific crowding and targeted bi- or multi-molecular interactions that can influence the free energy landscape of a protein. This heterogeneity can be so dramatic that an average over all free energy landscapes would not accurately represent the complexity of protein dynamics inside the cell (Fig. 7.2D).

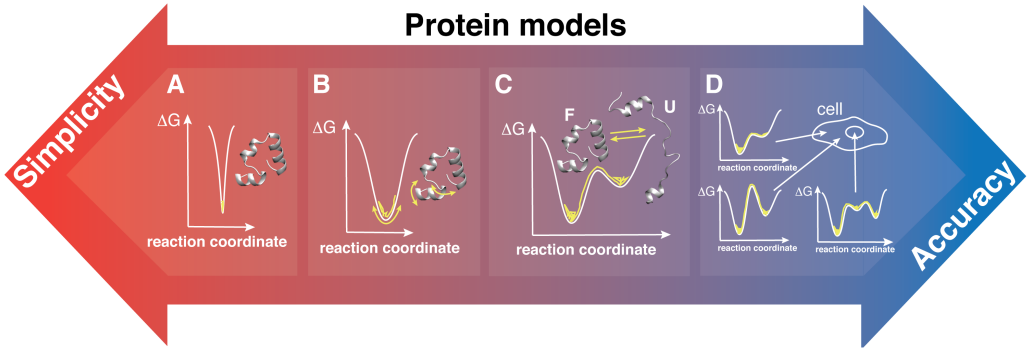


Figure 7.2: Models of proteins as a function of simplicity and accuracy of each model. (A) A rigid crystallographic state, which is commonly used to visualize a protein. If this state was the only one that actually existed in reality, the free energy surface for the protein would look like a narrow well with the minimum in free energy corresponding to the crystallographic structure. (B) A protein under solution conditions that strongly favor the folded state would largely only populate the native conformation, but thermal fluctuations would cause it to "breathe" deviating slightly from the crystallographic state. This behavior can be conceptualized as a broadening of the free energy well in (A). (C) A protein under solvent conditions where the folded state is preferred but the conformational space can be explored beyond the folded well to populate unfolded structures. An extra well is added on the free energy surface for the unfolded state with a barrier separating the two states. (D) The same protein can exhibit a multitude of stabilities and folding mechanisms in a heterogeneous environment of the living cell. Such locally specified parameters as viscosity, pH, ionic strength and crowding can influence the exact shape of the free energy landscape of the protein.

Thus, the example of protein folding is convenient for the purpose of illustrating the utility and limitations of simple models in science because a wide range of complexity vs. accuracy space is available to explore. In this activity we create models of a 10-amino-acid protein whose structure approximates that of Alpha-Synuclein.[282] These models were explored using mechanical prototypes and simple computer simulations.

7.3 Mechanical prototype

Students use mechanical prototypes of a 10-amino-acid protein whose folded structure approximates that of Alpha-Synuclein in that it is half unstructured, half alpha-helical (Figure 7.3A). The prototype for the unfolded states of this model protein consists of 9 paper clips connected together by flexible loops. The prototype for the folded protein is also made of 9 paper clips, but with 4 of the joints soldered so that half the protein is fixed in a rigid zig-zag, a proxy for a two-dimensional alpha helix. Each joint between the 5-cm-long connectors represents an amino acid. In order to investigate the two-dimensional conformational space that these two prototypical proteins would explore, the end-to-end distance is used as a reaction coordinate. End-to-end distance is a convenient variable to use in this case because it can be quickly measured with a ruler and it is also a common reaction coordinate in real biophysical studies of proteins where dyes are attached to each end of the protein and the distance-dependent energy transfer between them provides the signal.[283] Students throw the protein prototypes on the desk several times and measure the end-to-end distances for each resulting conformation. The results expected for this protein are contrary to those expected for the globular proteins most commonly used as instructional examples, as this proteins average end-to-end distance will increase when the protein is folded.

The students tabulate the resulting end-to-end distances and build a histogram of the number of occurrences vs. end-to-end distance. The center of mass of the histogram can be calculated as another way of reducing the dimensionality of the results further to compare the histograms that were obtained for the folded protein and the unfolded protein. This exercise is not only useful to discuss the physical concepts behind protein folding including entropy, enthalpy and the Gibbs free energy, but also to introduce students

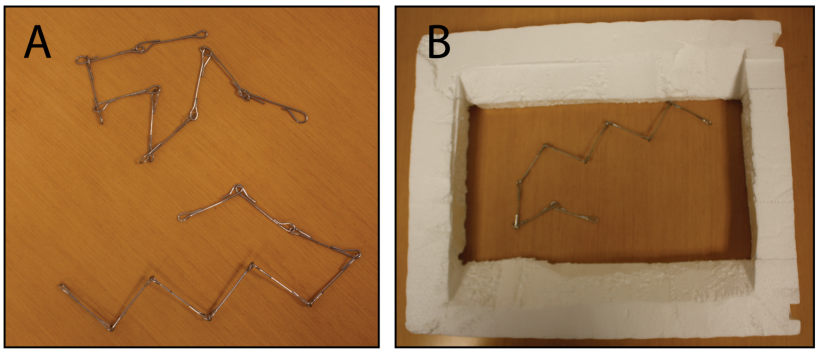


Figure 7.3: A two-dimensional mechanical prototype of a 10 amino-acid protein made with paper clips. The folded protein contains a rigid zigzag created by soldering 4 joints on one side of the chain. (B) A Styrofoam box is used to confine the protein prototype to a rectangular area to approximate cellular crowding.

to such statistical/mathematical tools and concepts as dimensionality reduction, histograms, weighted average, and various sources of error in physical measurements. Students also consider protein folding inside a cell. The concentration of macromolecules confined to the volume of a typical E. coli cell can be on the order of 300-400 g/L.[284] These molecules are sterically hindered from sampling the conformations that occupy large extended areas of space, i.e. the conformations that have large end-to-end distance. Students use a simple model for this crowding: a rectangular boundary made out of Styrofoam (Figure 7.3B). The Styrofoam frame serves as a spatial constraint preventing the protein from extending and thus decreasing its conformational entropy.

7.4 Computer simulations

Computer simulations are becoming increasingly more important in the biological sciences as the processors become faster and cheaper and algorithms get better at approximating the behavior of biological systems. Our approach synergistically uses the fundamental concepts of physical chemistry and scientific modeling, while also drawing on the tools from modern technology using custom software to simulate the experiment. There is ample evidence that computer simulations can have a positive effect on enhancing instruction, suggesting that technology is rapidly becoming a vital part of science

education.[285, 286] Simulations have been used previously to teach macromolecular structure[287] and kinetics,[288] however, to our knowledge, this is the first report where a computer simulation is specifically designed to reflect both structural and kinetic aspects of the mechanical model in question.

Students use a simple computer program (Figure 7.4) to simulate various structures that the paper clip protein model could form, to systematically analyze these structures, and to compare experimental results with simulation. The computer model mimics the mechanical one and computes end-to-end distances in the same way. Because the program can run through thousands of iterations in a matter of seconds, students can generate histograms that are much smoother than the ones obtained manually, which provides an opportunity to elaborate on the importance of statistical sampling. Students also compare the results obtained by using the mechanical model and the values that are generated by the software. The students are asked to discuss the discrepancies between the two results and determine the possible reasons why the computer simulation does not perfectly describe the behavior of the paper clip protein. In addition to the end-to-end distance statistics, the program also computes an equilibrium constant and kinetic time constant based on student input on the probability of the protein to switch between the two states. The simulation provides an opportunity to reinforce the idea that equilibria and kinetics are governed by statistical principles. This concept can lead to follow-up discussions about many thermodynamic and kinetic ideas appropriate to the course level, which may include the relationships between ΔG, activation barriers, and equilibrium distributions based on free energy surfaces.

7.5 Summary

We piloted this activity with students at both the high school and undergraduate introductory levels. Students in both settings enjoyed the activity and were surprised to learn that the same kinetics and thermodynamics concepts they had been applying to small molecules could also be applied to proteins, which they had mostly learned about in descriptive terms in biology classes.

Some students believe that models must be rigorously complete in order to be useful. Although this activity provides students with an opportunity to

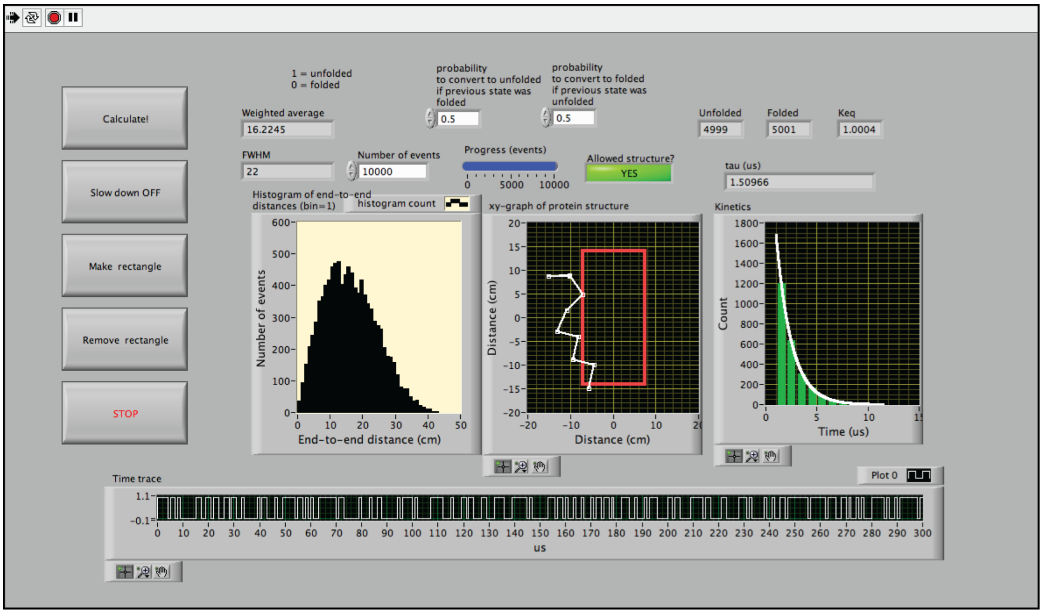


Figure 7.4: User interface of the computer program used to model the behavior of the prototypical protein. As inputs the program requires the number of samplings, the probability to switch to the unfolded state if the current state is folded and the conjugate probability going in the other direction. The outputs include a series of folded and/or unfolded protein conformations, a histogram of their end-to-end distances, the center of mass of the histogram, a plot of dwell times in each state as well as the equilibrium constant and the folding rate.

use an increasingly complex model to sample different equilibrium states for a protein, it should be emphasized that even the most sophisticated versions of models are incomplete descriptions, yet they still provide utility. This activity will guide students to a better understanding of the relationship between the statistics of a two-state system and its free energy surface and should be used as a complement to a larger ongoing discussion about the way in which models are used in science.

APPENDIX A

SUPPLEMENTARY MATERIAL FOR "REDUCING LAMBDA REPRESSOR TO THE CORE"

A.1 Explicit solvent replica-exchange molecular dynamics of λ_{6-85}*

The initial λ_{6-85} structure was taken from the 1LMB X-ray structure and modified to include the five mutations D14A/Y22W/Q33Y/G46A/G48A as described in [195]. The protein was modeled using the AMBER parm96 force field and solvated in a cubic box of TIP3P water with a minimum solute-wall distance of 25 Å. All ionizable groups, including N and C termini, were set to their expected protonation state at *p*H 7, resulting in a net protein charge of +1 that was neutralized by the addition of a Cl^- counterion. The final system consisted of 1258 protein atoms, one counterion, and 21904 water molecules, for a total of 66728 atoms.

After being set up, the system was subjected to a protocol of minimization and constant pressure equilibration. First, high-energy van der Waals contacts and bond lengths were repaired by 2000 steps of unconstrained steepest descent minimization using a 8.0 Å nonbonded cutoff without long range corrections. A further 2000 steps of steepest descent minimization with bonds constrained to their equilibrium lengths and a 9 Å nonbonded cutoff with particle mesh Ewald (PME) electrostatics prepared the system for constant pressure equilibration. Molecular dynamics (MD) simulations used a 2 fs timestep with all bonds constrained to within 10–8 Å of their equilibrium values by SHAKE and SETTLE. Nonbonded interactions used a 9 Å cutoff along with a longrange van der Waals correction. Long-range electrostatics were handled with PME using a grid spacing of 0.98 Å fourth order interpo-

* This Appendix is partially reproduced from the Supplementary Information from Maxim B. Prigozhin, Krishnarjun Sarkar, Dennis Law, William C. Swope, Martin Gruebele and Jed Pitera, Journal of Physical Chemistry B, 115(9), 2090–2096, 2011, DOI: 10.1021/jp110175x

lation, and an α value of 0.30768 $Å^{-1}$. All molecular dynamics simulations were carried out using the PMEMD module of AMBER8.

The volume of the simulated system was equilibrated by 10 ns of molecular dynamics at 310 K. During this equilibration phase, the temperature was controlled by an Andersen thermostat with a collision period of 1 ps. The volume was adjusted to its equilibrium value with a Berendsen weak-coupling scheme with an external pressure of 1 atm, an isothermal compressibility of 44.6×10^{-6} bar^{-1}, and a relaxation time constant of 5 ps. The final density after equilibration was 0.9786 g/cm^3, corresponding to a box edge length of 88.027 Å. While this system size is not sufficient to encompass the fully extended λ_{6-85} ($\sim$275 Å), a random coil of 80 amino acids has a mean end to end distance of approximately 34 Å.

The final conformation from the volume equilibration was used to start constant volume (NVT) replica exchange molecular dynamics (REMD) at 256 exponentially spaced temperatures ranging from 250 to 625 K. This spacing of approximately 1.1 K/replica at 300 K was chosen to produce an estimated acceptance ratio of 0.2 (actual 0.4). All replicas were started in the same folded, equilibrated conformation. During the first 1 ns of REMD stochastic exchanges were carried out every 10 ps, which was extended to 100 ps after that point. Temperature was controlled by Andersen collisions every 5 ps for the first 11 ns of REMD, with collisions every 100 ps thereafter. Protein coordinates were saved every 10 ps and full coordinates saved every 100 ps for later analysis. The primary REMD run was carried out for 179 ns/replica, for a total aggregate simulation time of 45.8 μs.

We plan to describe a more extensive analysis of this REMD simulation in a subsequent publication. However, the preliminary data (not shown) revealed higher stabilities for helices 1 and 4 as compared to other secondary structural elements of the protein. Interhelical contacts between these two helices were also preserved at temperatures up to approximately 400 K, where helices 2, 3 and 5 are all denatured. These observations led us to the present work, where we attempted to minimize the lambda repressor protein by retaining helices 1 and 4 and removing all other secondary structural elements.

Table A.1: Average radii of gyration and uncertainty (in Å) over the last 1.25 ns of each folded simulation

Variant name	300 K	350 K	400 K	450 K
h1aGSGh2a	11.4 ± 0.04	13.6 ± 0.71	15.7 ± 0.89	18.4 ± 0.79
h1aGSGh2b	11.2 ± 0.04	11.3 ± 0.09	17.0 ± 0.68	26.9 ± 1.25
h1aGSGh2c	11.3 ± 0.05	13.4 ± 0.26	15.9 ± 1.23	21.4 ± 1.09
h1aGSGSGh2a	11.2 ± 0.04	11.4 ± 0.07	16.3 ± 0.78	18.8 ± 0.71
h1aGSGSGh2b	11.3 ± 0.03	12.1 ± 0.14	14.6 ± 0.25	19.4 ± 1.24
h1aGSGSGh2c	11.2 ± 0.07	12.3 ± 0.12	19.8 ± 1.32	23.0 ± 0.43
h1aGSGSh2a	10.9 ± 0.06	11.8 ± 0.15	19.5 ± 1.94	22.7 ± 1.01
h1aGSGSh2b	11.2 ± 0.01	12.2 ± 0.19	17.3 ± 0.52	20.7 ± 1.31
h1aGSGSh2c	10.7 ± 0.05	11.5 ± 0.13	18.8 ± 1.55	17.6 ± 1.46
h1aWTh2a	11.8 ± 0.08	12.1 ± 0.06	20.5 ± 0.98	26.7 ± 1.52
h1aWTh2b	11.6 ± 0.11	12.0 ± 0.18	16.9 ± 1.12	24.0 ± 1.10
h1bGSGh2a	11.4 ± 0.11	12.6 ± 0.53	15.9 ± 0.82	23.2 ± 0.94
h1bGSGh2b	11.1 ± 0.05	13.2 ± 0.66	15.8 ± 0.97	21.8 ± 1.06
h1bGSGh2c	11.1 ± 0.04	12.2 ± 0.60	17.1 ± 0.90	22.7 ± 1.16
h1bGSGSGh2a	11.0 ± 0.09	11.7 ± 0.23	19.2 ± 1.25	24.1 ± 1.13
h1bGSGSGh2b	10.9 ± 0.09	12.1 ± 0.18	17.9 ± 1.59	22.3 ± 1.26
h1bGSGSGh2c	11.0 ± 0.05	11.7 ± 0.20	14.8 ± 0.85	20.7 ± 0.44
h1bGSGSh2a	11.1 ± 0.08	12.0 ± 0.30	14.8 ± 0.75	20.7 ± 0.72
h1bGSGSh2b	11.2 ± 0.13	12.5 ± 0.29	15.7 ± 0.62	20.1 ± 1.18
h1bGSGSh2c	10.9 ± 0.18	12.9 ± 0.46	15.9 ± 0.56	20.2 ± 1.62
h1bSGSGh2a	11.3 ± 0.09	13.7 ± 0.62	15.9 ± 0.72	18.9 ± 0.67
h1bSGSGh2b	11.2 ± 0.04	11.8 ± 0.43	16.6 ± 0.62	25.2 ± 1.31
h1bSGSGh2c	11.3 ± 0.04	13.1 ± 0.24	16.0 ± 1.00	21.8 ± 1.01
h1cGSGh2a	11.3 ± 0.08	12.4 ± 0.40	18.5 ± 1.55	23.6 ± 1.53
h1cGSGh2b	11.0 ± 0.06	11.4 ± 0.16	17.4 ± 0.94	24.9 ± 1.56
h1cGSGh2c	10.8 ± 0.08	11.7 ± 0.22	16.0 ± 0.64	21.7 ± 1.46
h1cGSGSGh2a	10.7 ± 0.12	12.0 ± 0.18	17.0 ± 0.95	21.1 ±1.40
h1cGSGSGh2b	11.0 ± 0.15	11.8 ± 0.12	17.4 ± 1.16	24.5 ±1.61
h1cGSGSGh2c	11.0 ± 0.09	11.8 ± 0.09	17.0 ± 1.28	23.0 ±1.65
h1cGSGSh2a	10.9 ± 0.06	11.6 ± 0.10	16.5 ± 1.06	19.6 ±0.58
h1cGSGSh2b	10.8 ± 0.06	11.6 ± 0.16	15.8 ± 0.66	18.6 ±0.66
h1cGSGSh2c	10.8 ± 0.06	11.4 ± 0.17	15.8 ± 0.36	19.2 ± 1.02
Average	11.1	12.1	16.9	21.8

Table A.2: Average radii of gyration and uncertainty (in Å) over the last 1.25 ns of each unfolded simulation

Variant name	300 K	350 K	400 K	450 K
h1aGSGh2a	11.7 ± 0.12	10.9 ± 0.03	18.1 ± 0.67	20.0 ± 1.21
h1aGSGh2b	11.2 ± 0.03	11.9 ± 0.09	14.6 ± 0.40	23.2 ± 1.22
h1aGSGh2c	11.0 ± 0.05	13.7 ± 0.23	17.0 ± 1.03	22.8 ± 1.42
h1aGSGSGh2a	12.0 ± 0.09	11.7 ± 0.09	13.3 ± 0.41	21.5 ± 1.71
h1aGSGSGh2b	11.3 ± 0.06	16.3 ± 0.13	16.4 ± 0.77	20.4 ± 1.26
h1aGSGSGh2c	10.9 ± 0.15	15.4 ± 0.66	15.5 ± 0.79	21.5 ± 0.77
h1aGSGSh2a	12.2 ± 0.07	12.7 ± 0.26	17.8 ± 0.24	25.1 ± 1.16
h1aGSGSh2b	10.9 ± 0.02	12.0 ± 0.09	18.3 ± 0.67	19.3 ± 1.05
h1aGSGSh2c	10.1 ± 0.05	12.7 ± 0.29	13.9 ± 0.30	20.7 ± 0.96
h1aWTh2a	13.9 ± 0.04	13.9 ± 0.54	20.8 ± 1.12	24.1 ± 1.49
h1aWTh2b	12.6 ± 0.03	12.8 ± 0.14	16.4 ± 1.01	22.7 ± 0.87
h1bGSGh2a	10.5 ± 0.07	13.1 ± 0.14	15.1 ± 0.44	17.3 ± 0.48
h1bGSGh2b	10.8 ± 0.02	12.9 ± 0.09	15.6 ± 0.39	19.4 ± 1.10
h1bGSGh2c	10.4 ± 0.03	13.3 ± 0.18	18.1 ± 1.21	20.2 ± 0.89
h1bGSGSGh2a	12.3 ± 0.09	12.0 ± 0.18	20.3 ± 1.63	21.6 ± 0.97
h1bGSGSGh2b	10.5 ± 0.02	12.4 ± 0.23	14.9 ± 0.25	22.5 ± 1.06
h1bGSGSGh2c	11.0 ± 0.05	10.5 ± 0.09	14.5 ± 0.54	21.4 ± 1.38
h1bGSGSh2a	10.8 ± 0.10	12.1 ± 0.30	14.8 ± 0.42	25.9 ± 1.98
h1bGSGSh2b	11.9 ± 0.03	13.1 ± 0.17	17.9 ± 1.42	23.8 ± 1.19
h1bGSGSh2c	11.2 ± 0.09	18.3 ± 0.32	17.3 ± 1.23	20.0 ± 1.27
h1bSGSGh2a	12.2 ± 0.18	11.5 ± 0.07	15.4 ± 0.58	21.6 ± 0.96
h1bSGSGh2b	12.3 ± 0.14	11.8 ± 0.25	18.7 ± 0.46	20.1 ± 1.32
h1bSGSGh2c	11.5 ± 0.08	12.0 ± 0.20	14.3 ± 0.63	22.0 ± 0.88
h1cGSGh2a	11.3 ± 0.04	10.6 ± 0.12	16.8 ± 0.73	21.0 ± 1.40
h1cGSGh2b	11.4 ± 0.08	10.9 ± 0.04	17.7 ± 0.35	23.8 ± 0.68
h1cGSGh2c	11.2 ± 0.06	12.4 ± 0.22	16.9 ± 0.62	23.1 ± 1.54
h1cGSGSGh2a	10.6 ± 0.03	12.9 ± 0.25	18.4 ± 1.61	23.3 ± 1.53
h1cGSGSGh2b	12.1 ± 0.10	11.6 ± 0.24	16.0 ± 0.67	18.9 ± 1.20
h1cGSGSGh2c	10.1 ± 0.03	11.8 ± 0.09	16.3 ± 1.21	21.1 ± 0.79
h1cGSGSh2a	10.7 ± 0.06	10.7 ± 0.08	15.6 ± 1.06	21.4 ± 0.69
h1cGSGSh2b	10.6 ± 0.05	10.8 ± 0.08	13.5 ± 0.31	22.7 ± 1.20
h1cGSGSh2c	10.2 ± 0.04	11.0 ± 0.11	16.0 ± 0.72	21.5 ± 1.14
Average	11.3	12.5	16.4	21.7

Table A.3: Sequences for each lambda repressor variant

Variant name	Residues	Atoms	Amino acid sequence
h1aGSGh4a	50	800	LTQEQLEAARRLKAIWEKKKNELGLSGSGFNGINALNAYNAALLAKILKV
h1aGSGh4b	56	888	LTQEQLEAARRLKAIWEKKKNELGLSGSGFNGINALNAYNAALLAKILKVSVEEFS
h1aGSGh4c	47	750	LTQEQLEAARRLKAIWEKKKNELGLSGSGGINALNAYNAALLAKILK
h1aGSGSh4a	51	811	LTQEQLEAARRLKAIWEKKKNELGLSGSGSFNGINALNAYNAALLAKILKV
h1aGSGSh4b	57	899	LTQEQLEAARRLKAIWEKKKNELGLSGSGSFNGINALNAYNAALLAKILKVSVEEFS
h1aGSGSh4c	48	761	LTQEQLEAARRLKAIWEKKKNELGLSGSGSGINALNAYNAALLAKILK
h1aGSGSGh4a	52	818	LTQEQLEAARRLKAIWEKKKNELGLSGSGSGFNGINALNAYNAALLAKILKV
h1aGSGSGh4b	58	906	LTQEQLEAARRLKAIWEKKKNELGLSGSGSGFNGINALNAYNAALLAKILKVSVEEFS
h1aGSGSGh4c	49	768	LTQEQLEAARRLKAIWEKKKNELGLSGSGSGGINALNAYNAALLAKILK
h1aWTh4a	65	1023	LTQEQLEAARRLKAIWEKKKNELGLSYESVADKMGMGQSAVAALFNGINALNAYN AALLAKILKV
h1aWTh4b	71	1111	LTQEQLEAARRLKAIWEKKKNELGLSYESVADKMGMGQSAVAALFNGINALNAYN AALLAKILKVSVEEFS
h1bGSGh4a	51	821	LTQEQLEAARRLKAIWEKKKNELGLSYGSGFNGINALNAYNAALLAKILKV
h1bGSGh4b	57	909	LTQEQLEAARRLKAIWEKKKNELGLSYGSGFNGINALNAYNAALLAKILKVSVEEFS
h1bGSGh4c	48	771	LTQEQLEAARRLKAIWEKKKNELGLSYGSGGINALNAYNAALLAKILK
h1bGSGSh4a	52	832	LTQEQLEAARRLKAIWEKKKNELGLSYGSGSFNGINALNAYNAALLAKILKV
h1bGSGSh4b	58	920	LTQEQLEAARRLKAIWEKKKNELGLSYGSGSFNGINALNAYNAALLAKILKVSVEEFS
h1bGSGSh4c	49	782	LTQEQLEAARRLKAIWEKKKNELGLSYGSGSGINALNAYNAALLAKILK
h1bGSGSGh4a	53	839	LTQEQLEAARRLKAIWEKKKNELGLSYGSGSGFNGINALNAYNAALLAKILKV
h1bGSGSGh4b	59	927	LTQEQLEAARRLKAIWEKKKNELGLSYGSGSGFNGINALNAYNAALLAKILKVSVEEFS
h1bGSGSGh4c	50	789	LTQEQLEAARRLKAIWEKKKNELGLSYGSGSGGINALNAYNAALLAKILK
h1cGSGh4a	48	770	LTQEQLEAARRLKAIWEKKKNELGGSGFNGINALNAYNAALLAKILKV
h1cGSGh4b	54	858	LTQEQLEAARRLKAIWEKKKNELGGSGFNGINALNAYNAALLAKILKVSVEEFS
h1cGSGh4c	45	720	LTQEQLEAARRLKAIWEKKKNELGGSGGINALNAYNAALLAKILK
h1cGSGSh4a	49	781	LTQEQLEAARRLKAIWEKKKNELGGSGSFNGINALNAYNAALLAKILKV
h1cGSGSh4b	55	869	LTQEQLEAARRLKAIWEKKKNELGGSGSFNGINALNAYNAALLAKILKVSVEEFS
h1cGSGSh4c	46	731	LTQEQLEAARRLKAIWEKKKNELGGSGSGINALNAYNAALLAKILK
h1cGSGSGh4a	50	788	LTQEQLEAARRLKAIWEKKKNELGGSGSGFNGINALNAYNAALLAKILKV
h1cGSGSGh4b	56	876	LTQEQLEAARRLKAIWEKKKNELGGSGSGFNGINALNAYNAALLAKILKVSVEEFS
h1cGSGSGh4c	47	738	LTQEQLEAARRLKAIWEKKKNELGGSGSGGINALNAYNAALLAKILK

APPENDIX B

SUPPLEMENTARY MATERIAL FOR "THE FAST AND THE SLOW: FOLDING AND TRAPPING OF λ_{6-85}"

B.1 Thermal titrations of λ_{6-85} mutants monitored by fluorescence and CD*

* This Appendix is partially reproduced from the Supplementary Information from Maxim B. Prigozhin and Martin Gruebele, Journal of the American Chemical Society, 133(48), 19338–19341, 2011, DOI: 10.1021/ja209073z

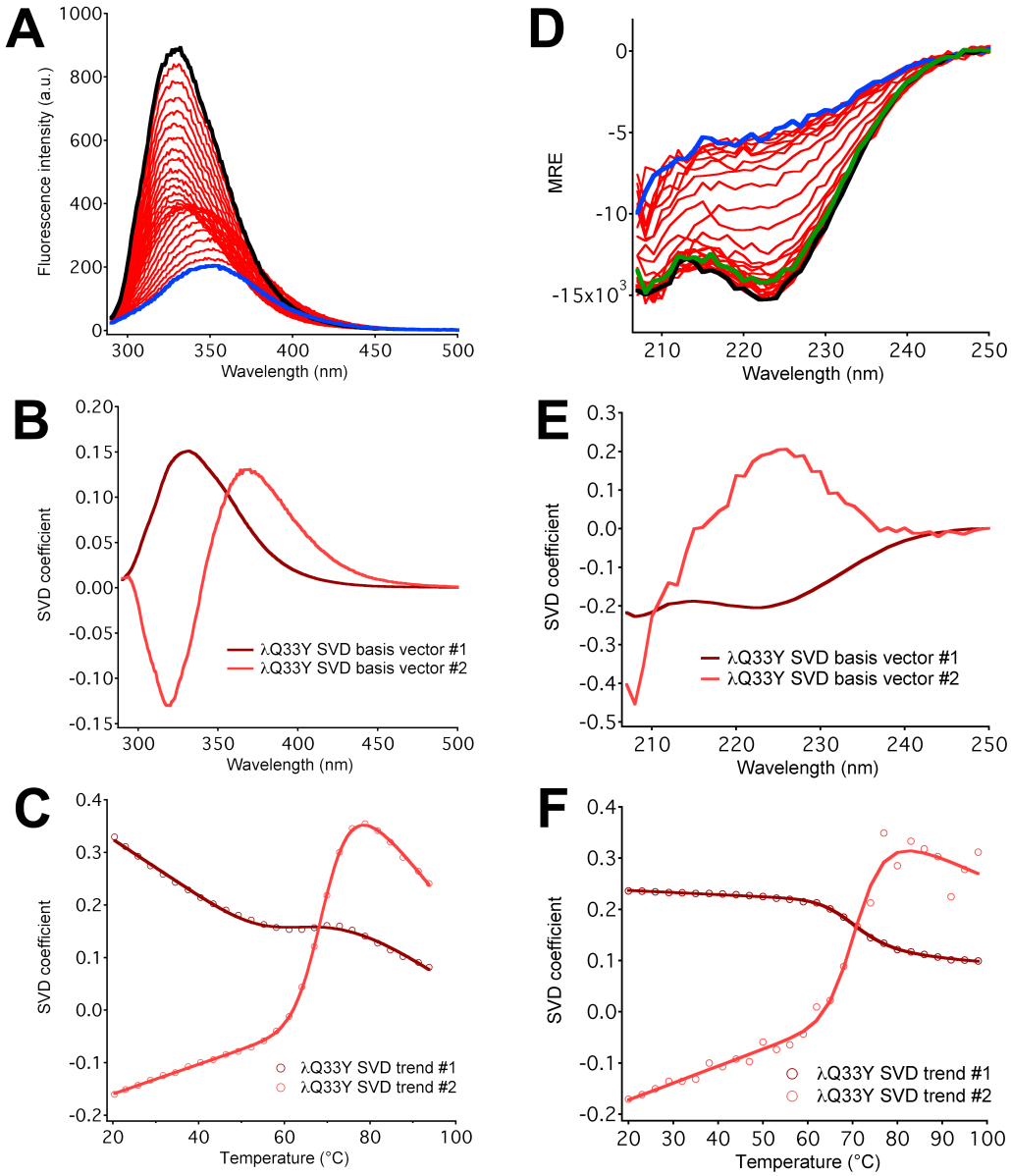


Figure B.1: Thermal titrations of λQ33Y monitored by fluorescence (A–C) and CD (D–F). (A) Spectra collected at ∼3 °C increments from 20 °C (black curve) to 94 °C (blue curve). Integrated intensity decreases with increasing temperature. (B) SVD basis vectors. (C) SVD trends with corresponding two-state thermodynamic fits. (D) Spectra collected at 3 °C increments from 20 °C (black curve) to 98 °C (blue curve). After the melt was done, the temperature was again decreased to 20 °C to show reversibility (green curve). (E) SVD basis vectors. (F) SVD trends with corresponding two-state thermodynamic fits.

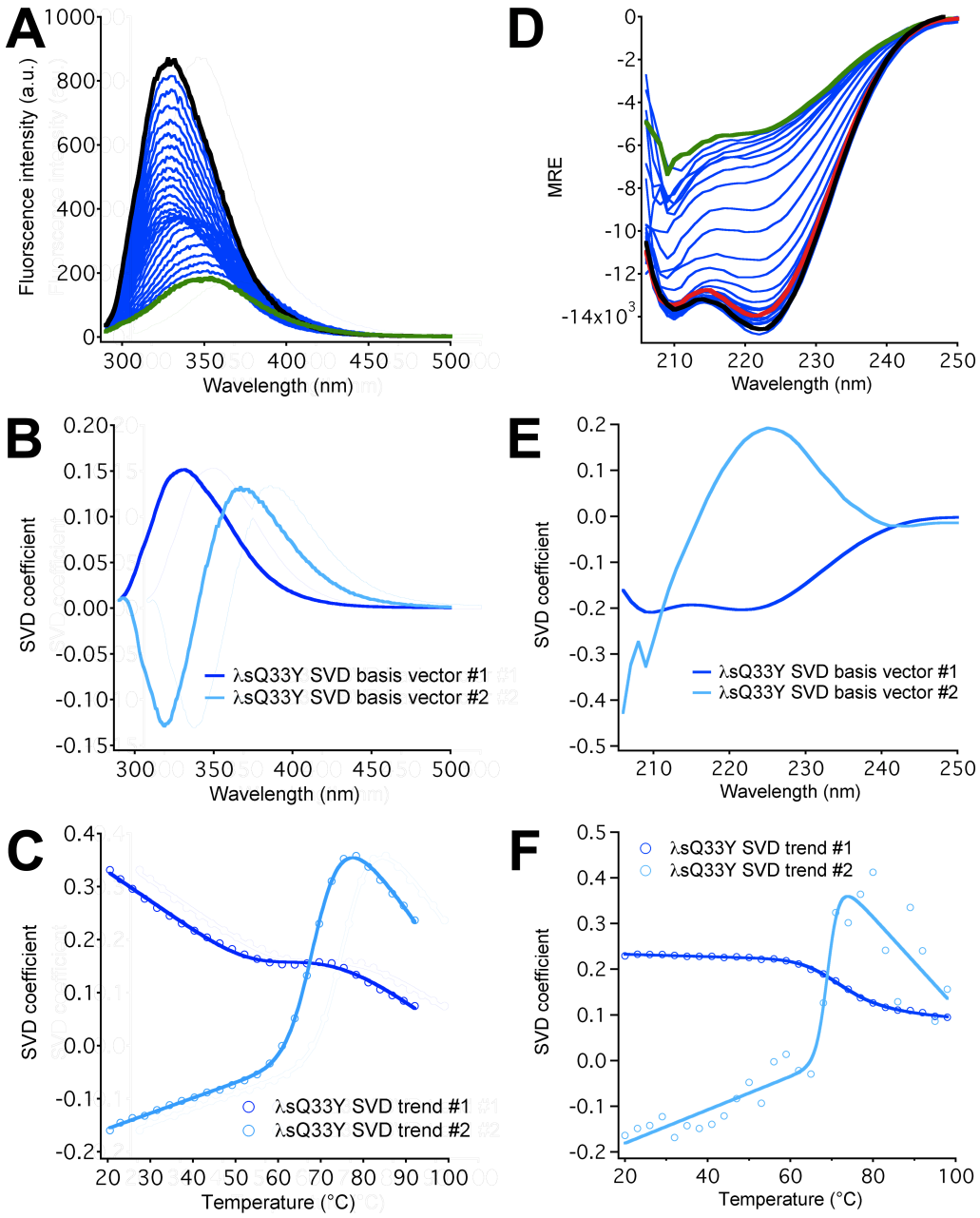


Figure B.2: Thermal titrations of λsQ33Y monitored by fluorescence (A–C) and CD (D–F). (A) Spectra collected at ∼3 °C increments from 20 °C (black curve) to 94 °C (green curve). Integrated intensity decreases with increasing temperature. (B) SVD basis vectors. (C) SVD trends with corresponding two-state thermodynamic fits. (D) Spectra collected at 3 °C increments from 20 °C (black curve) to 98 °C (green curve). After the melt was done, the temperature was again decreased to 20 °C to show reversibility (red curve). (E) SVD basis vectors. (F) SVD trends with corresponding two-state thermodynamic fits.

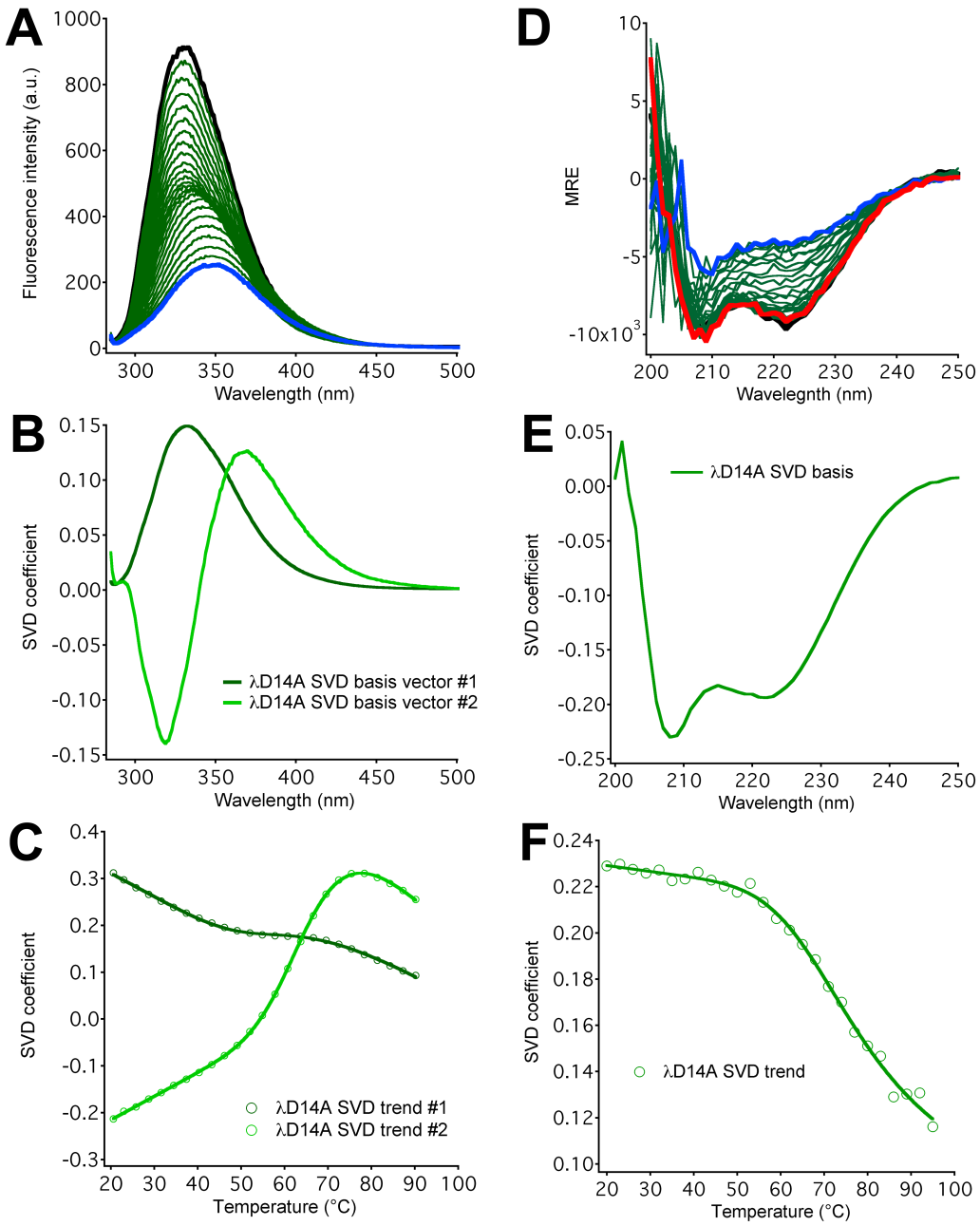


Figure B.3: Thermal titrations of λD14A monitored by fluorescence (A–C) and CD (D–F). (A) Spectra collected at ∼3 °C increments from 20 °C (black curve) to 94 °C (blue curve). Integrated intensity decreases with increasing temperature. (B) SVD basis vectors. (C) SVD trends with corresponding two-state thermodynamic fits. (D) Spectra collected at 3 °C increments from 20 °C (black curve) to 98 °C (blue curve). After the melt was done, the temperature was again decreased to 20 °C to show reversibility (red curve). (E) SVD basis vectors. (F) SVD trends with corresponding two-state thermodynamic fits.

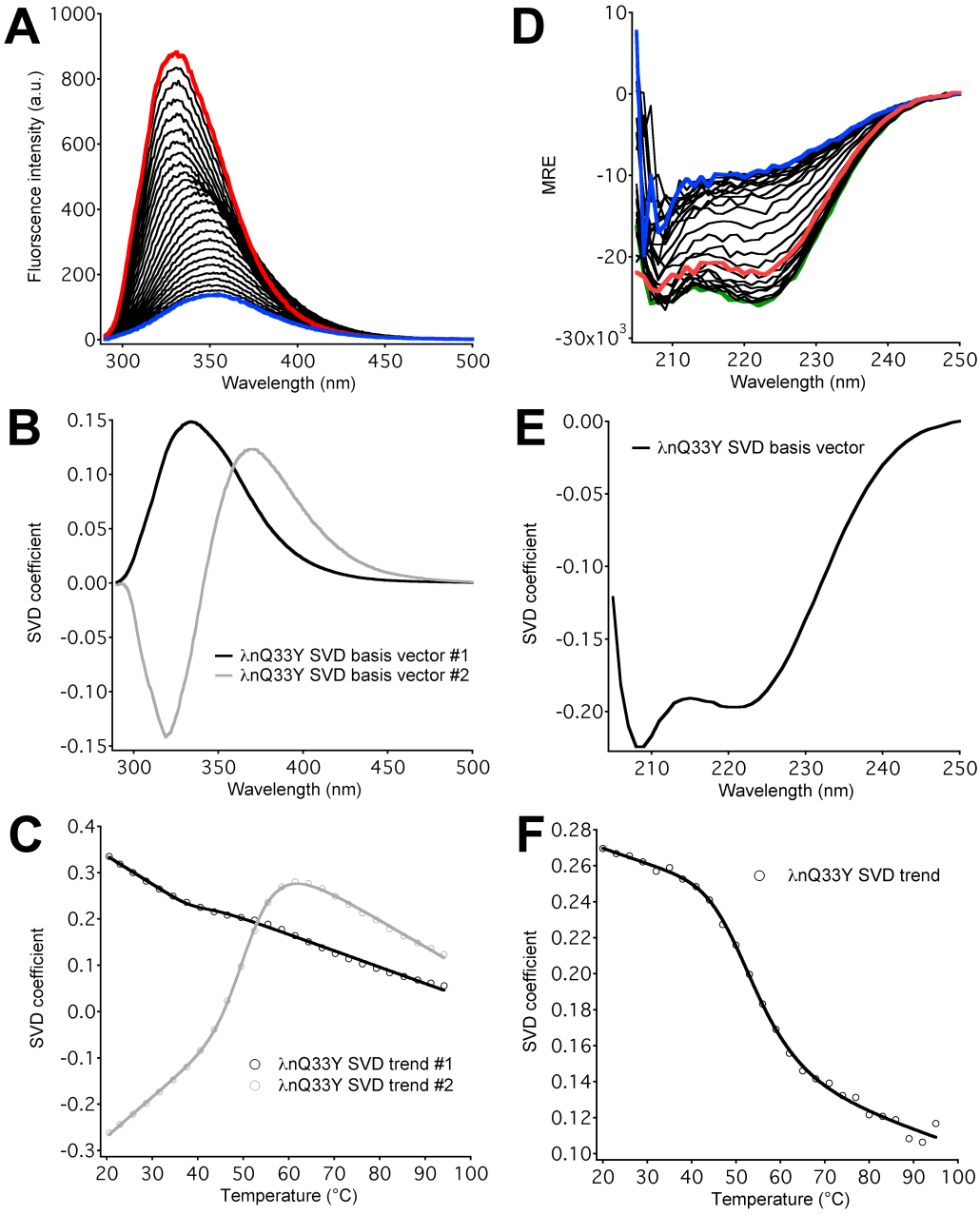


Figure B.4: Thermal titrations of λnQ33Y monitored by fluorescence (A–C) and CD (D–F). (A) Spectra collected at ∼3 °C increments from 20 °C (red curve) to 94 °C (blue curve). Integrated intensity decreases with increasing temperature. (B) SVD basis vectors. (C) SVD trends with corresponding two-state thermodynamic fits. (D) Spectra collected at 3 °C increments from 20 °C (red curve) to 98 °C (blue curve). After the melt was done, the temperature was again decreased to 20 °C to show reversibility (green curve). (E) SVD basis vectors. (F) SVD trends with corresponding two-state thermodynamic fits.

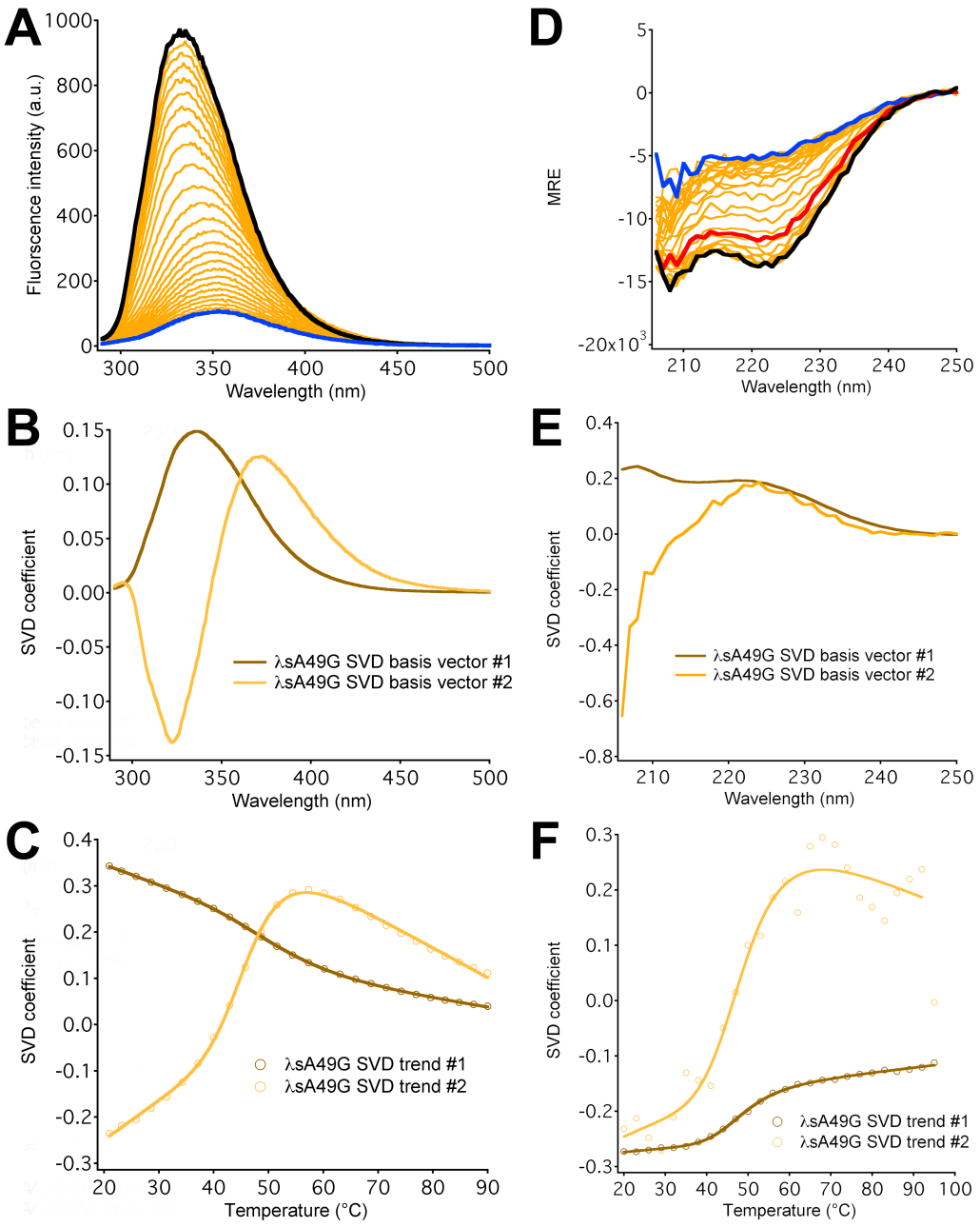


Figure B.5: Thermal titrations of λsA49G monitored by fluorescence (A–C) and CD (D–F). (A) Spectra collected at ∼3 °C increments from 20 °C (black curve) to 94 °C (blue curve). Integrated intensity decreases with increasing temperature. (B) SVD basis vectors. (C) SVD trends with corresponding two-state thermodynamic fits. (D) Spectra collected at 3 °C increments from 20 °C (black curve) to 98 °C (blue curve). After the melt was done, the temperature was again decreased to 20 °C to show reversibility (red curve). (E) SVD basis vectors. (F) SVD trends with corresponding two-state thermodynamic fits.

Table B.1: Thermodynamic midpoints for unfolding of λ-repressor mutants as measured by tryptophan fluorescence and CD spectroscopy. Here SVD 1 represents the thermodynamic midpoints determined by the decrease in the intensity and SVD 2 represents the thermodynamic midpoints determined by the wavelenght shift of the W22 fluorescence spectrum or CD spectrum of λ-repressor mutants. SVD 2 could not be extracted from λD14A and λnQ33Y data because the shift of CD spectrum was not large for these mutants. The reported values are based on changes in the intensity of the CD spectrum at 222 nm.

	Temperature (°C)				
	Fluorescence		CD		
Mutant	SVD 1	SVD 2	SVD 1	SVD 2	Reported values[31]
λD14A	63 ± 3	68.0 ± 0.3	77 ± 14	–	73.5
λQ33Y	68 ± 4	70.3 ± 0.1	71.0 ± 0.5	72 ± 1	71
λsQ33Y	68 ± 4	69.8 ± 0.2	73 ± 1	69.3 ± 0.9	61.5
λnQ33Y	40 ± 2	52.6 ± 0.2	55 ± 1	–	54.5
λsA49G	55.1 ± 0.5	47.6 ± 0.3	48.7 ± 0.5	50 ± 3	47

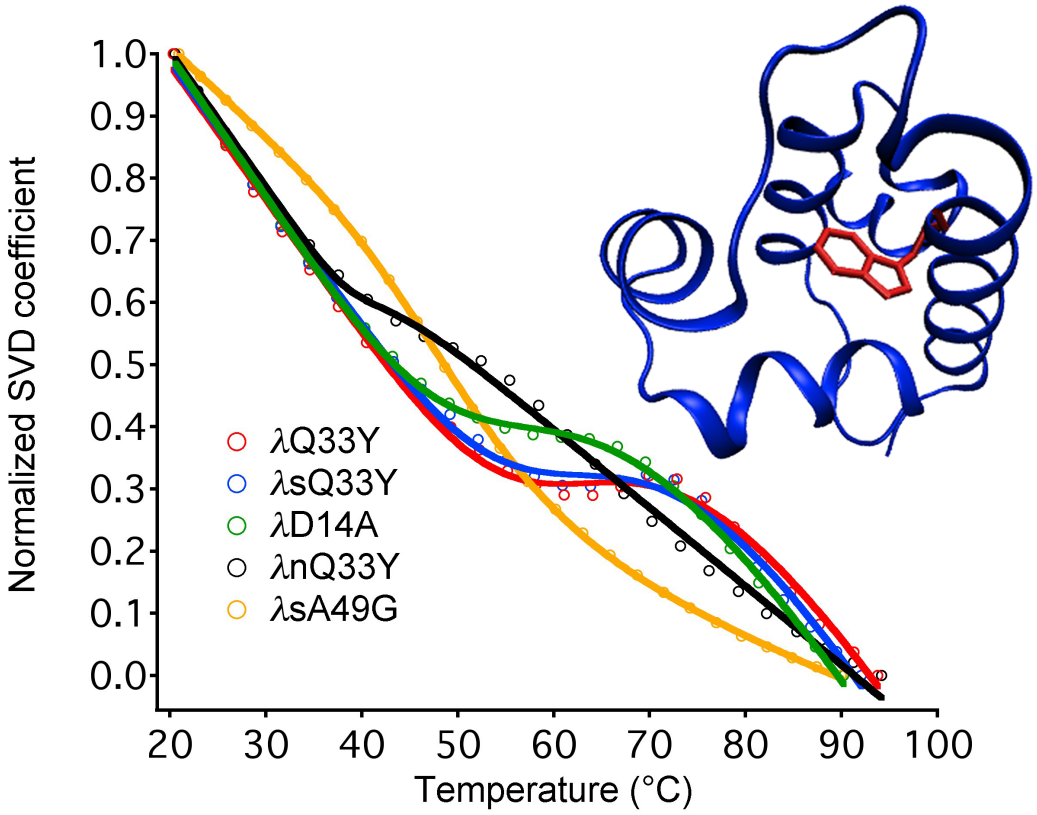


Figure B.6: Thermal denaturation of λ-repressor mutants. SVD shown here represents the decrease in the intensity of the W22 fluorescence spectrum. The inset shows the crystal structure of λQ33Y. Note the qualitatively different behavior of the λs49G mutant (yellow), which lacks the Q33Y mutation. All other mutants show a relative increase of fluorescence upon denaturation, superimposed on an overall decrease of tryptophan fluorescence intensity at higher temperature, as observed also for tryptophan itself.

B.2 Temperature jump kinetics of λ_{6-85} mutants spanning 5 ms

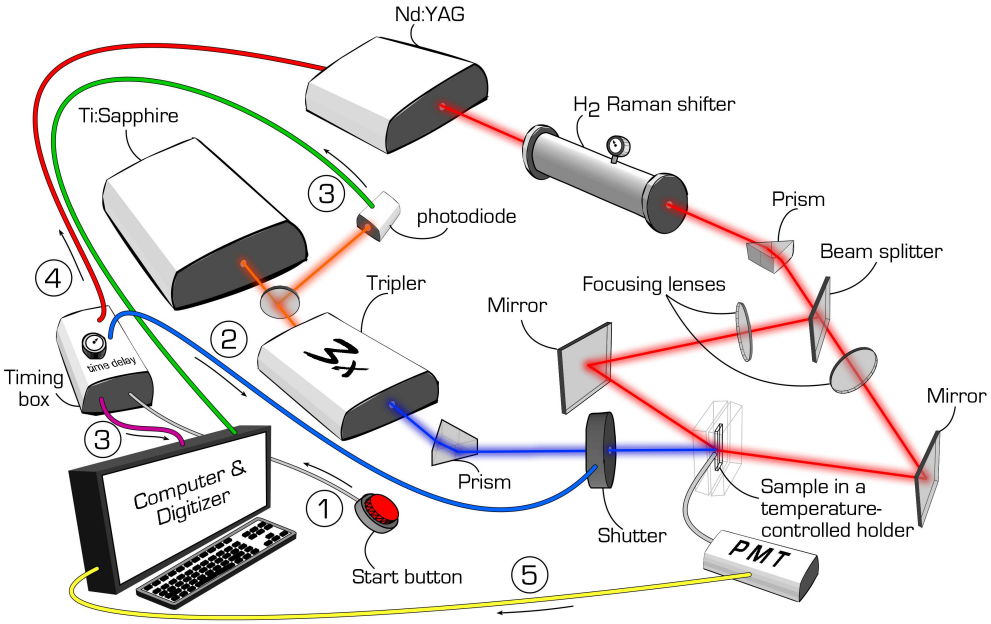


Figure B.7: Optical components of the nanosecond laser temperature jump instrument.

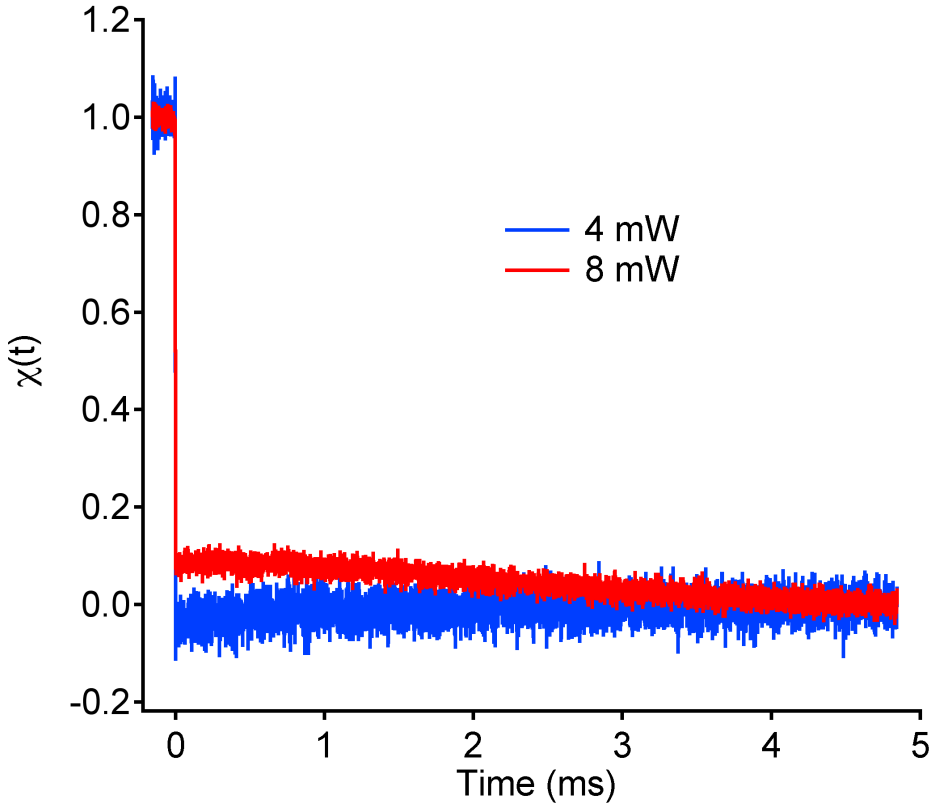


Figure B.8: 5 °C temperature jumps of 300 μM *L*-tryptophan solution to the final temperature of 45 °C at incident power of the UV beam from Ti:Sapphire laser equal to 8 mW and 4 mW. The traces are normalized such that the average of the first and last 100 data points of $\chi(t)$ are set to 1 and 0, respectively. At 4 mW excitation power, fluorescence lifetime of tryptophan increases slightly after the jump due to slow recooling of the sample to the equilibrium temperature. At 8 mW, the cooling trend is overpowered by the opposing effect of irreversible photobleaching, which reduces the fluorescence lifetime of tryptophan over time.

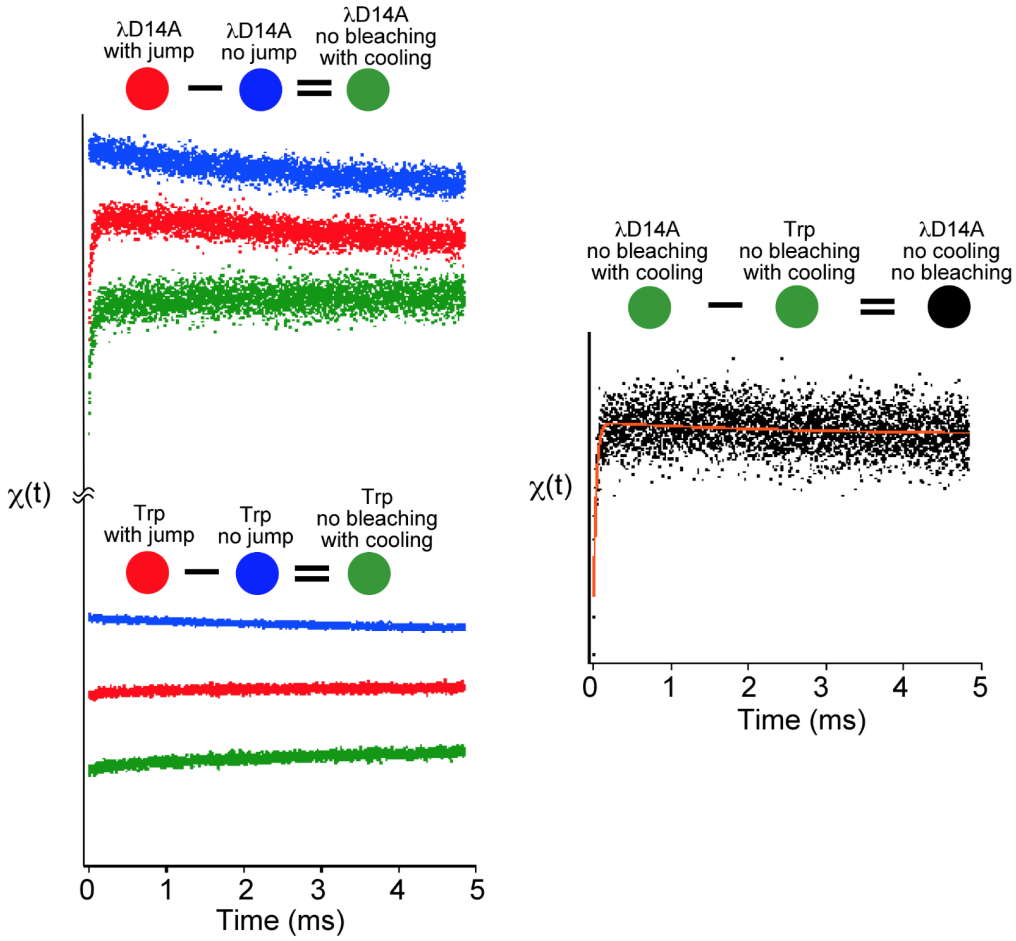


Figure B.9: Correction of kinetics data for the effects of cooling and bleaching. The overall trends observed remained the same even without this correction. Top left: Kinetic trace of λD14A to the final temperature of 60 °C is shown in red. A similar set of data is shown for λD14A at 60 °C but without a jump (hence no folding phase) in blue. Kinetic trace in red contains the effects of cooling and bleaching along with the folding signature of λD14A but the trace in blue only contains the effect of λD14A bleaching at 60 °C, which is subtracted out from the red trace to yield the green time trajectory. The latter then contains the effects of cooling and folding of λD14A at 60 °C. Bottom left: The same procedure was followed to account for the bleaching effect in *L*-tryptophan at 60 °C. Center right: Assuming cooling of λD14A and Trp at 60 °C after a T-jump are similar we subtract the Trp time trace that contains only the cooling effect from that of λD14A that has both the cooling and the folding signature. The resulting time trace (black) can then be fitted to a double exponential (orange curve).

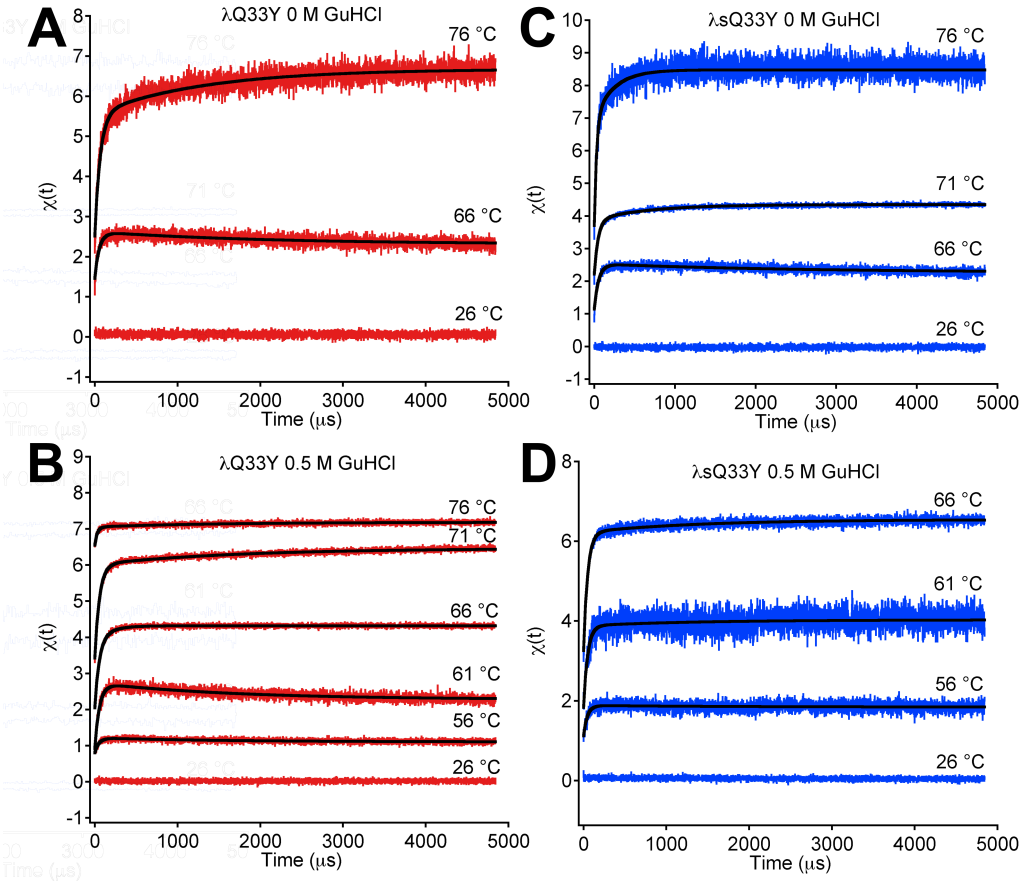


Figure B.10: Temperature jumps of the more stable mutants of λ-repressor: λQ33Y (A and B) and λsQ33Y (C and D) at 0 M (A and C) and 0.5 M (B and D) GuHCl. Black curves represent the bi-exponential fits of the data. The temperature jumps occurred at $t = 0$. Temperatures after the jump are indicated above the corresponding time traces.

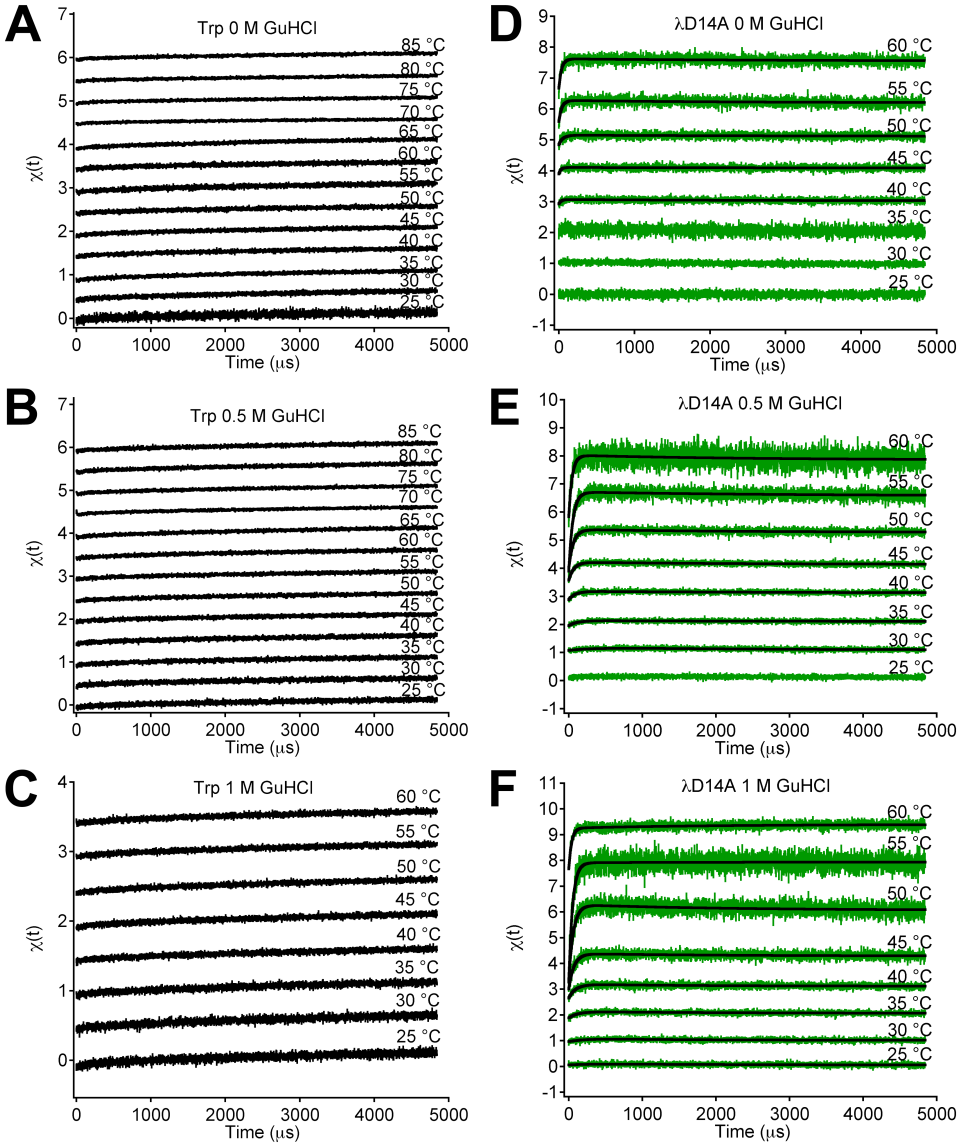


Figure B.11: Temperature jumps of *L*-tryptophan (A–C) and λD14A (D–F), a mutant of intermediate thermal stability, at 0 M (A and D), 0.5 M (B and E), and 1 M (C and F) GuHCl. Black curves represent the bi-exponential fits of the data. The temperature jumps occurred at $t = 0$. Temperatures after the jump are indicated above the corresponding time traces.

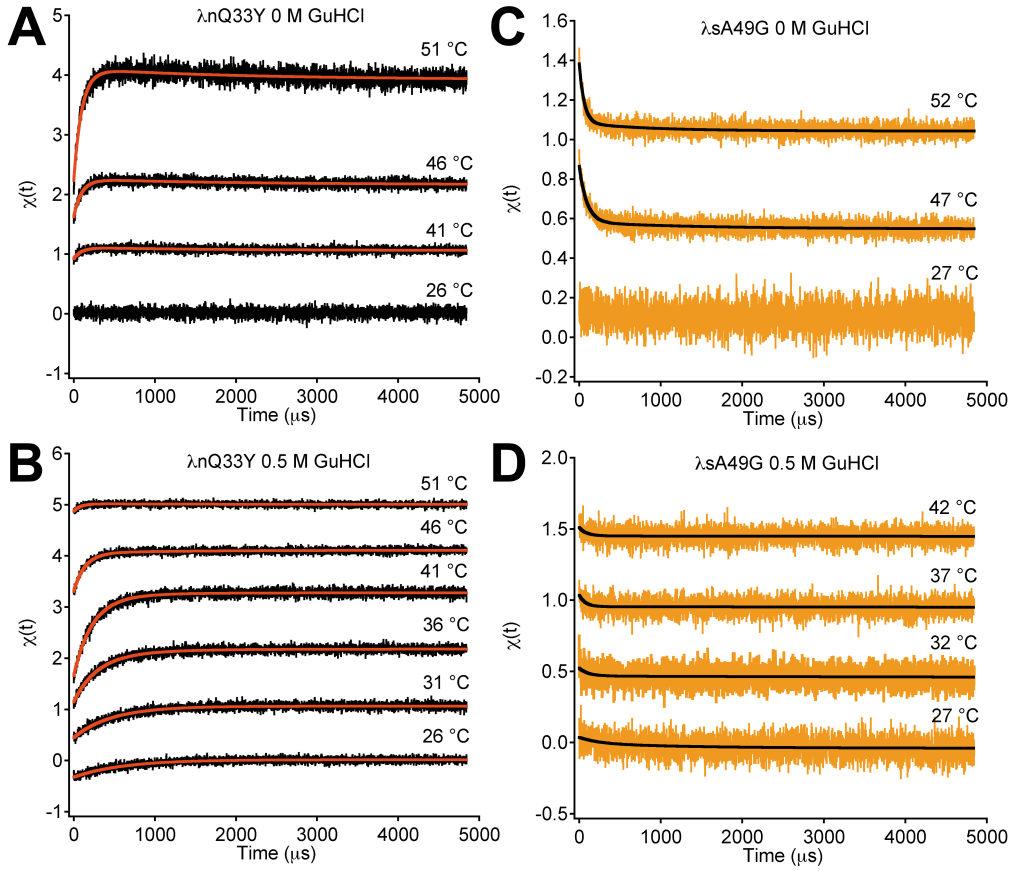


Figure B.12: Temperature jumps of the less stable mutants of λ-repressor: λnQ33Y (A and B) and λsA49G (C and D) at 0 M (A and C) and 0.5 M (B and D) GuHCl. Orange (λnQ33Y) and black (λsA49G) curves represent the bi-exponential fits of the data. The temperature jumps occurred at $t = 0$. Temperatures after the jump are indicated above the corresponding time traces.

B.3 Global model fit

We fitted the data with several 3 state models. Data at several T-jump temperatures were fitted simultaneously by each 3 state model distinguishing N, E and T, but not T and T′, nor accounting for the ∼1 to 2 μs phase attributed to downhill folding of λQ33Y or λD14A. We do not distinguish T and T′ in the fit because we cannot distinguish these states experimentally, although they are very distinct in the simulations of Bowman *et al*. We do not fit the ultrafast downhill folding phase because it has been investigated in detail previously, and here we are concerned with the interconversion of N, E and T: it is only of interest that N and E interconvert rapidly (either two-state over a low barrier, or downhill). We used a kinetic global master equation program previously described in detail by Sarkar *et al*.[289] to fit RNA folding data. Double exponential decays such as in Figure 3.3A of the main paper between 0.01 and 5 ms can be accounted by a variety of master equations with different connectivity. Figure B.13 summarizes the schemes that produced satisfactory fits of the data.

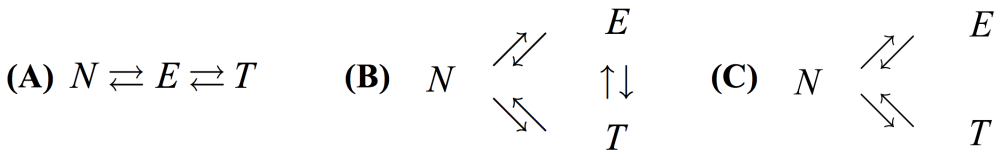


Figure B.13: Three-state kinetic schemes that were compatible with the experimental data regardless of realistic tryptophan lifetimes of the states N, E or T, or whether they are consistent with the computational results from Bowman *et al*. Of the models shown, only (A) and (B) produced reasonable relative tryptophan lifetimes (smaller = more quenched for N than for extensively unfolded state, T either quenched or similar to E). Model (A) is also consistent with the simulation results and with the notion that the native state can not directly convert to a compact beta sheet state without first going through an extensively unfolded state U⇔E.

Note that the first point of the low temperature data in Figure 3.3A corresponds to downhill folding, which was not accounted for by the global model and hence this data point was not included in the fit. As expected, downhill folding disappeared at high T.

It can be seen from the free energies that the extensively unfolded state becomes rapidly more stable than N when the temperature is increased, the

Table B.2: Thermodynamic parameters fitted or held fixed for Scheme S1(A) fitted to the λQ33Y data in Figure 3.3A (blue curves, main paper).

T, K	ΔG(N), kJ/mole	ΔG(E), kJ/mole	ΔG(T), kJ/mole	$\Delta G^{\dagger}$(N–E), kJ/mole	$\Delta G^{\dagger}$(E–T), kJ/mole	$\Delta G^{\dagger}$(N–T), kJ/mole
333	0	11.6	2.1	16.7	25.2	∞
339	0	6	0.7	16.7	25.2	∞
343	0	2.3	–0.2	16.7	25.2	∞
349	0	–3.3	–1.6	16.7	25.2	∞

trap T more slowly so. At the highest temperatures, the extensively unfolded state is lowest in free energy, at the lowest temperatures, the native state is lowest in free energy.

The free energies were fitted as (standard deviation errors indicated, none if fixed):

$\Delta G(E) = -0.929 \pm 0.023$ (T–345.5 ± 0.13)

$\Delta G(T) = -0.23 \pm 0.04$ (T–342)

$\Delta G^{\dagger}(E–T) = 25.10 \pm 0.35$

and the relative signal baselines of the three states were (no temperature dependence was required for fitting):

N: 1 (fixed)

E: 1.30 ± 0.023

T: 0.974 ± 0.002

APPENDIX C

SUPPLEMENTARY MATERIAL FOR "MISPLACED HELIX SLOWS DOWN ULTRAFAST PRESSURE-JUMP PROTEIN FOLDING"

C.1 Kinetic data*

NATA and protein pressure drops starting at 1200 bar were collected on our P-jump instrument. Repeated P-jumps of NATA showed that the dead-time of the P-jump depends on pressure change, and at 1200 bar a 2–3 μs jump is obtained.

C.2 Guanidine titrations

Titrations with GuHCl buffered at *p*H 7 were performed to find a suitable denaturation midpoint where the λ*YA mutant would be poised to unfold at room temperature. As can be seen from the SVD trends in panel C, for the intensity (blue) and wavelength shift (red) basis functions in panel B, 2.4 M guanidine lies just below the unfolding transition, lowering the pressure denaturation midpoint from *ca.* 6 kbar (0 M GuHCl) to *ca.* 1.5 kbar (2.4 M GuHCl).

C.3 Temperature titrations

Temperature titrations were monitored by IR (main text) and CD (below). A Jasco-715 spectropolarimeter (Jasco Inc., Easton, MD) was used to measure the circular dichroism spectra of λ*YA. The spectra are averages of 3 spectra, which were collected at 100 nm/min each. The measurements were done

* This Appendix is partially reproduced from the Supplementary Information from Maxim B. Prigozhin, Yanxin Liu, Anna Jean Wirth, Shobhna Kapoor, Roland Winter, Klaus Schulten and Martin Gruebele, Proceedings of the National Academy of Sciences of the United States of America, 110(20), 8087–8092, 2011, DOI: 10.1073/pnas.1219163110

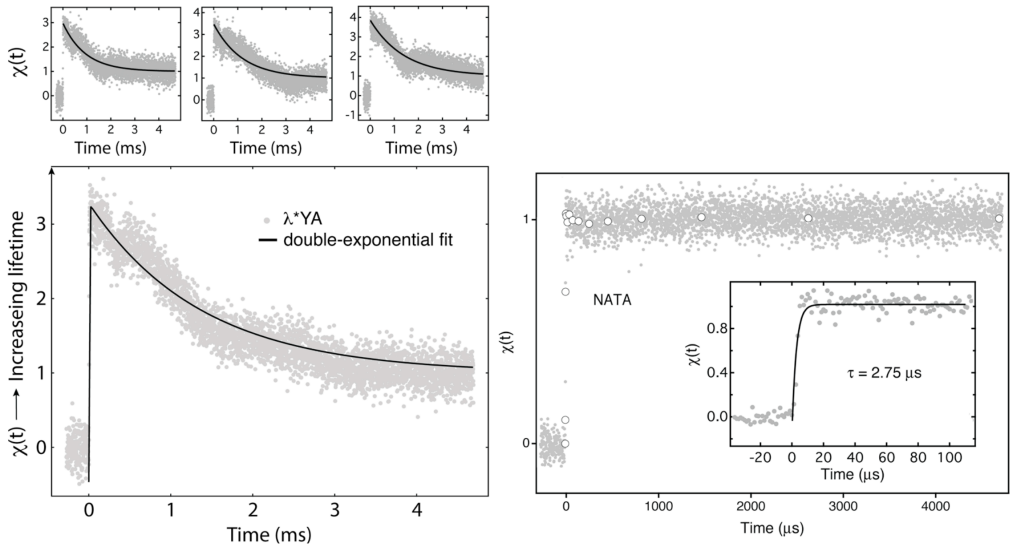


Figure C.1: The refolding kinetic data in Figure 4.3 of the main text is an average of 3 pressure drops of λ*YA in 2.4 M GuHCl. Here the data are analyzed with respect to the Trp fluorescence lifteime decay of the protein ($\chi = 0$ means the data is perfectly fitted by the decay f_1 before the jump, $\chi = 1$ means the data is perfectly fitted by the decay f_2 at 5 ms, other values of χ mean the data is fitted by a linear combination $(1 - \chi)f_1 + \chi f_2$. On the right the NATA P-jump (also analyzed with its own f_1 and f_2 functions) is highlighted near $t = 0$, showing the time resolution obtained for 1200 bar P-jumps, which is somewhat lower than for the 2400 bar P-jumps we reported previously (Dumont *et al.* reference in main text).

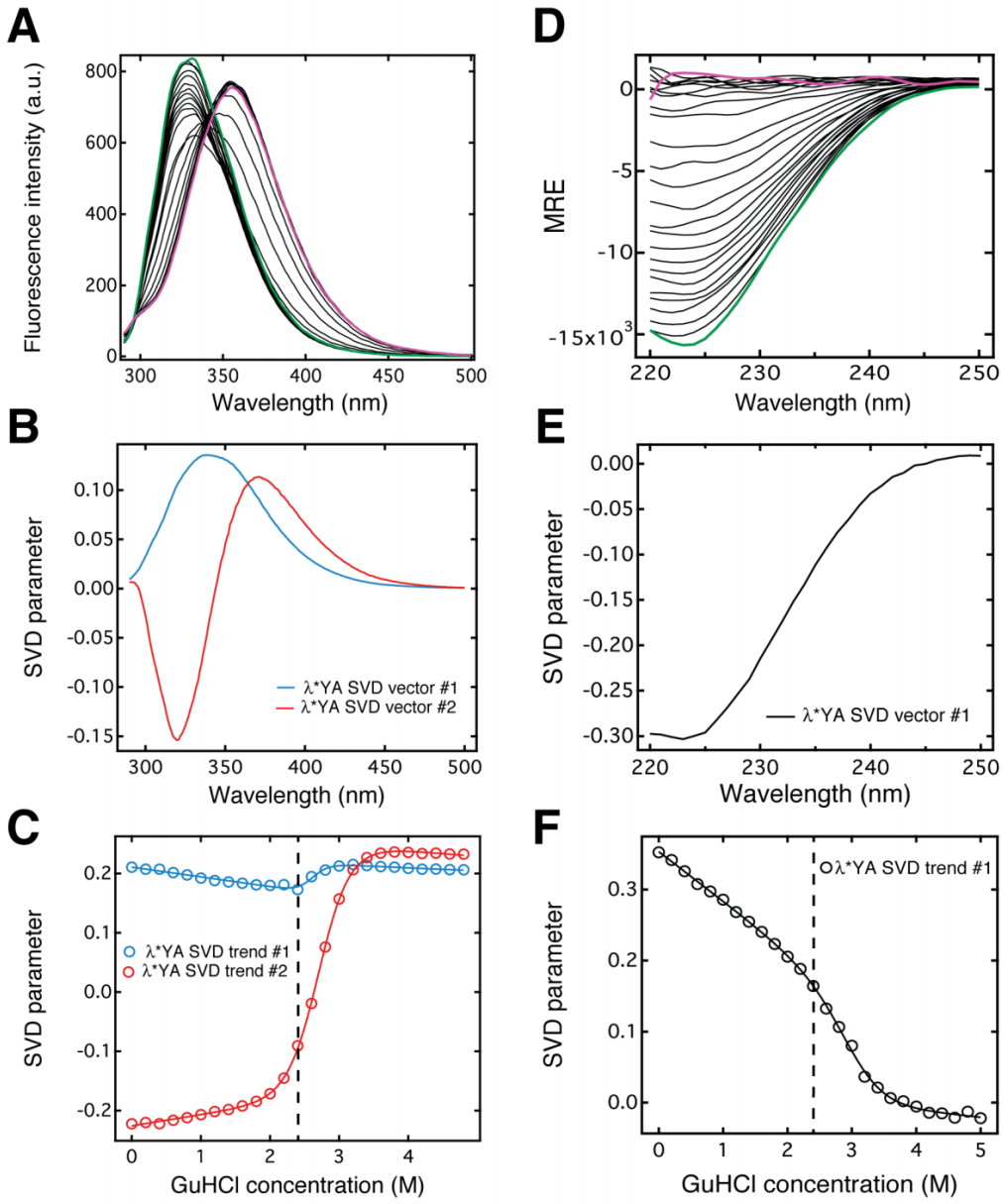


Figure C.2: GuHCl titration of λ*YA showing 2.4 M onset of denaturation. The titration was monitored by fluorescence (A–C, 5 μM sample) and CD (D–F, 25 μM sample). (A, D) Spectra collected at 0.2 M increments of GuHCl from 0 M (green curve) to 5 M (magenta curve). (B, E) SVD basis vectors. (C, F) SVD trends with corresponding two-state thermodynamic fits. Vertical dashed lines highlight the 2.4 M GuHCl point.

using a quartz cuvette (Starna Cells Inc., Atascadero, CA) with either 1 mm or 1 cm path length. We used singular value decomposition to analyze the data.

We found that a protein concentration of 2.5 μM produces a nearly (>90%) reversible denaturation curve, 50 μM is still >80% reversible, but 100 μM and above is nearly irreversible following temperature titration up to 95 °C. The IR spectra in the main text show that this correlates with formation of extended (β) structure, either in the monomer, or larger aggregates.

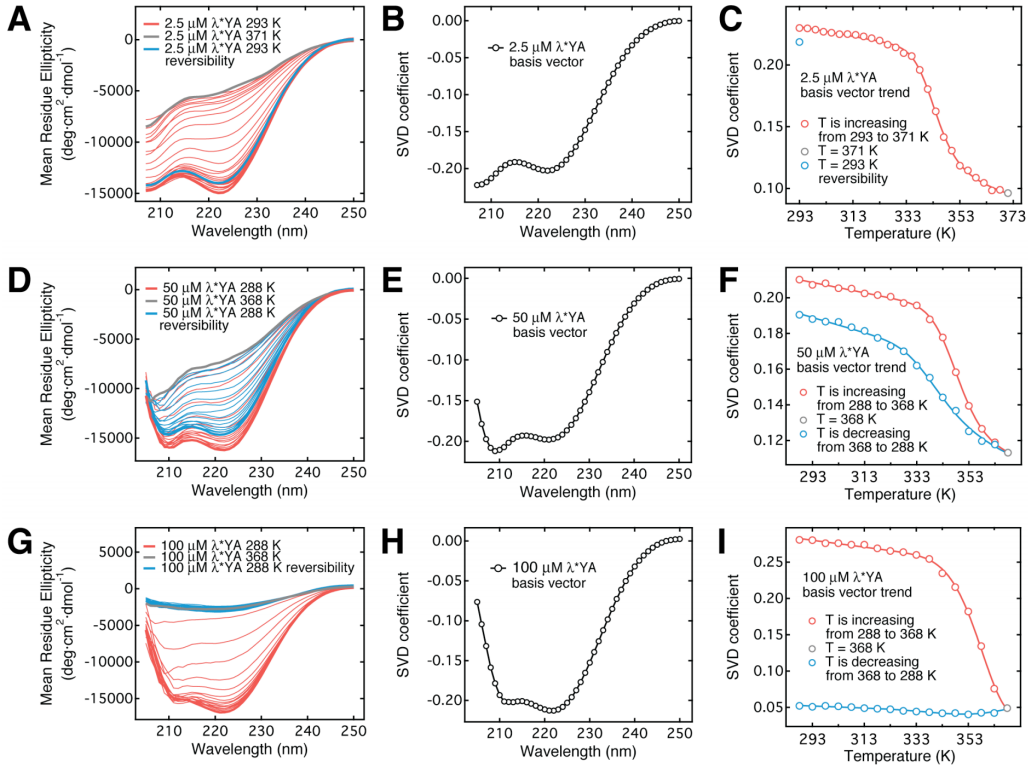


Figure C.3: Reversibility of λ*YA unfolding after a temperature denaturation at different protein concentrations as monitored by CD spectroscopy. (A–C) 2.5 μM, (D–F) 50 μM, (G–I) 100 μM. (A, D, G) CD spectra of λ*YA. Spectra in red follow the signal as the temperature increases, grey spectra represent the signal at the highest temperature, and spectra in blue show the signal as the temperature is decreasing back down. (B, E, H) The first SVD basis vectors for the corresponding data sets in A, D and G, respectively. (C, F, I) The first SVD basis vector trends color coded the same way as in panels A, D and G for the corresponding data sets in A, D and G, respectively.

C.4 Pressure titrations

Pressure titrations were used to establish changes in the amide I′ IR spectrum, Trp lifetime and spectrum upon denaturation. We found that NATA has a substantially longer lifetime than Trp in λ*YA. Nonetheless the NATA decay f_1 (at 1200 bar) and f_2 (1 bar) could be used to fit the protein data just as well as the corresponding fluorescence decays of the protein, by a simple two-state model

$$f = [1 - \chi(P)]f_1 + \chi(P)f_2.$$

The fit using NATA fluorescence decays as basis functions is shown in Figure 4.3 of the main text, the fit using protein decays is shown in Figure C.1. χ was also used to determine the fully unfolded spectrum in Figure 4.1A of the main text, by extrapolating χ from 1200 bar upwards until the function f was just about to go negative. This produces a conservative unfolded state spectrum, in the sense that the actual spectrum probably has higher intensity, and actual fraction unfolded is higher than we quote.

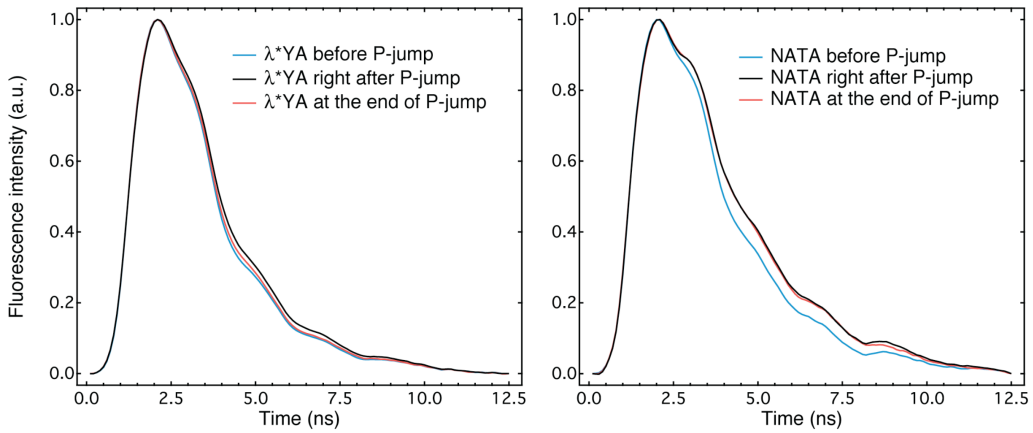


Figure C.4: Fluorescence decays of λ*YA (left) and NATA (right) before the pressure-jump (function f_1, P = 1.2 kbar, blue curve), right after the pressure-jump (P = 1 bar, black curve), and at the end of the time trace (function f_2, P = 1 bar, red curve). The decays shown here correspond to an average of 100 decays. Note that NATA has a significantly longer fluorescence lifetime than Trp in the protein, for this reason χ was shifted up by 3 units in for the protein lifetime data in Figure 4.1D.

A two-state model was also suficient to account for the IR pressure denaturation shown in Figure 4.4 of the main text. Figure C.5 above shows the error and reconstructed IR spectrum with largest error compared to the

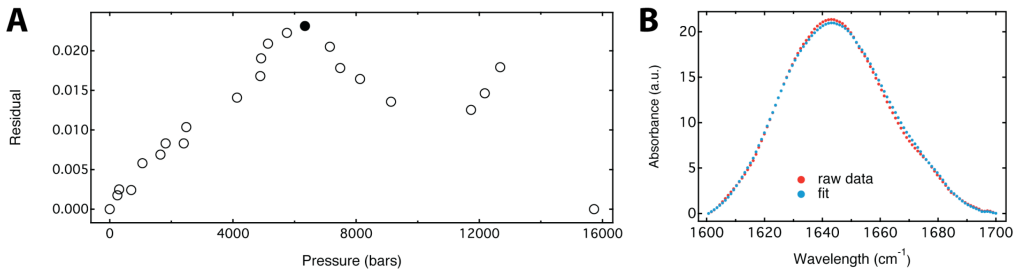


Figure C.5: Residual of fitting the IR-detected pressure denaturation to the basis functions in Figure 4.4A to obtain the denaturation curve in Figure 4.4B. (A) Residuals were obtained by subtracting the fitted spectra from the data, integrating over the wavelengths, and dividing the resulting value by the integral of the data spectrum. (B) The raw data as well as the fit for the largest residual value (solid black circle in (A)).

measured IR spectrum.

Unlike temperature denaturation, pressure denaturation was highly reversible up to twice the pressure used in our jumps. Nowhere did the denaturation-renaturation curves differ significantly from one another for either intensity or wavelength shift of the fluorescence spectrum.

Detecting pressure titrations in GuDCl buffer by infrared spectroscopy of the amide I′ band was difficult. The red edge of the protein band below 1650 cm^{-1} is obscured by strong absorption from the GuDCl reagent. At 2.4 M guanidinium, we were only able to tell that no β-sheet peak shows up at 1680 cm^{-1} upon pressure denaturation. At 0.5 M GuDCl (Figure C.7), it was possible to approximately deconvolve the guanidine and protein spectra by measuring guanidine along, protein alone, and both together as a function of pressure. The protein peak shifts towards 1640 cm^{-1} just as it did without guanidine, indicating transition from a helix to a more coil-like state. No evidence of a strong β-sheet peak at 1620 or 1680 cm^{-1} was seen, again showing that GuDCl does not change the pressure denaturation mechanism from alpha-coil to α-β/aggregate. From the combined fluorescence and IR data, we can say that unlike temperature denaturation, pressure denaturation is reversible and does not induce aggregation.

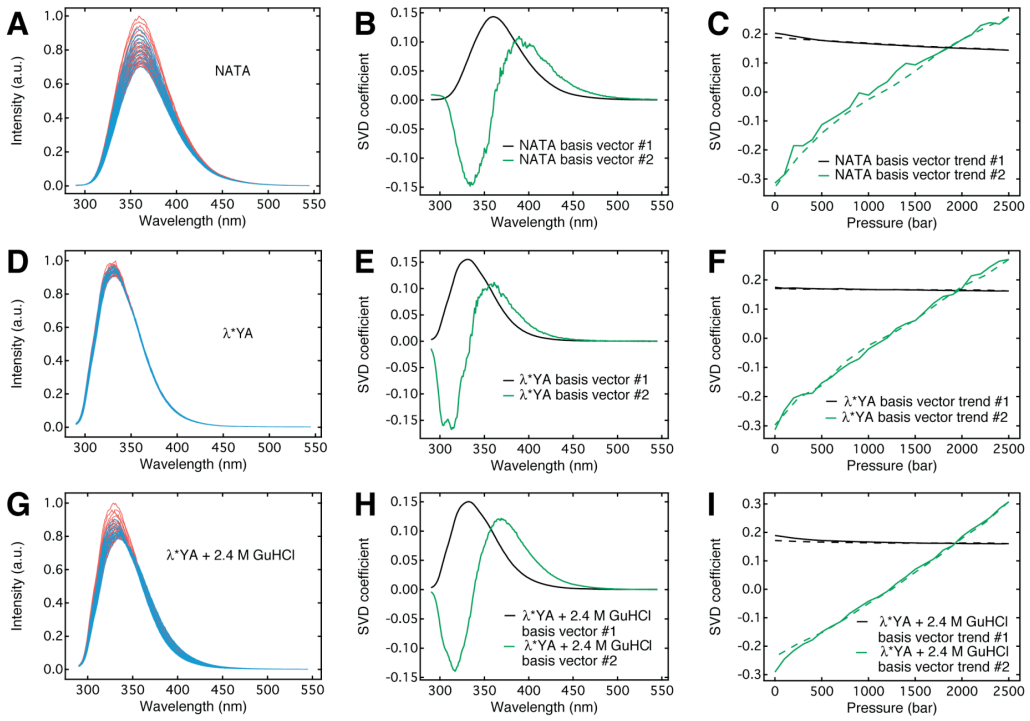


Figure C.6: Reversibility of NATA and λ*YA with and without GuHCl unfolding after a pressure denaturation as monitored by fluorescence spectroscopy. (A–C) NATA, (D–F) λ*YAwithout GuHCl, (G–I) λ*YA with GuHCl. (A, D, G) Fluorescence spectra of NATA, λ*YA without GuHCl, and λ*YA with GuHCl, respectively. Spectra in red follow the signal as the pressure increases, and spectra in blue show the signal as the pressure is decreasing back down. (B, E, H) The first (black) and the second (green) SVD basis vectors for the corresponding data sets in A, D and G, respectively. (C, F, I) The first (black) and the second (green) SVD basis vector trends for the corresponding data sets in A, D and G, respectively. The pressure increase is shown as a solid line and the pressure decrease is shown as a dashed line.

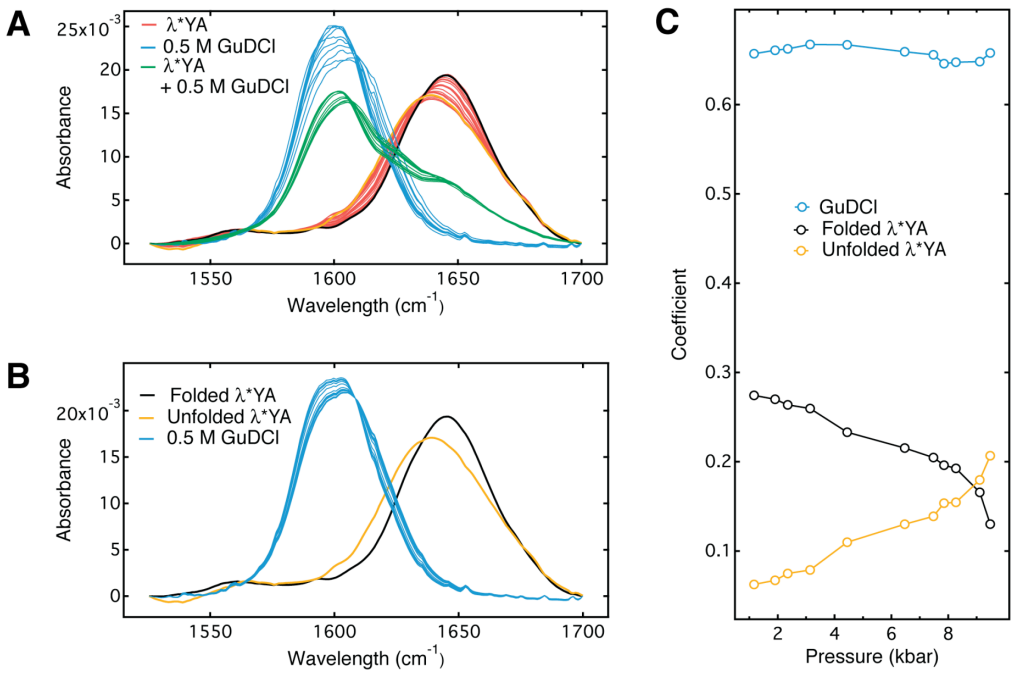


Figure C.7: FTIR spectroscopy of λ*YA with 0.5 M GuDCl. (A) In red: FTIR spectra of λ*YA shifting towards lower energy from 1 bar to 13.9 kbar without GuDCl. Basis vectors that were used for the folded (black) and unfolded (yellow) protein are also shown. In blue: FTIR spectra of 0.5 M GuDCl from shifting towards higher energy from 1 bar to 14.2 kbar. In green: FTIR spectra of λ*YA with 0.5 M GuDCl. (B) Basis spectra used for χ-analysis of the green spectra in panel A. Black and yellow spectra correspond to λ*YA at 1 bar and 13.9 kbar, respectively, as shown in panel A. The blue spectra were obtained by doing χ-analysis on the GuDCl spectra (blue spectra in panel A) separately, fitting the resulting $\chi(P)$ to a straight line and using that linear fit to arrive at GuDCl spectra at the pressures at which λ*YA with 0.5 M GuDCl (green spectra in panel A) was measured. (C) χ-coefficients for the GuDCl basis spectrum (blue), basis spectrum for the folded λ*YA (black), and the basis spectrum for the unfolded λ*YA (yellow). As expected, χ-coefficient for GuDCl stays flat, the folded population decreases, while the unfoled population increases as the pressure increases.

C.5 Molecular dynamics simulations

C.5.1 System setup

The structure for the λ-repressor fragment (λ*YA mutant) was obtained from the protein data bank (PDB code 1LMB)[290]. The native structure was placed in a box of 24,282 water molecules at 55 mM NaCl salinity and neutralized through extra ions using VMD[291]. The simulation box measured 91.1 Å each dimension at T = 325 K and P = 1 bar and contained 74,209 atoms including protein, water molecules and ions.

C.5.2 Simulation using NAMD

The upward temperature/pressure jump simulations, in which the protein unfolded to an initial state, were carried out using NAMD[292]. The simulated system was rst subjected to 6000 steps of conjugate gradient minimization and equilibrated for 200 ps with harmonic restraints applied to all protein heavy atoms. The simulations were then continued for 2 ns without restraints at constant pressure of 1 bar using a Nosé-Hoover Langevin piston barostat and at constant temperature of 329 K using a Langevin thermostat with damping constant of 5.0 ps^{-1}.

During the production runs, which include high P-T preparation of unfolded initial states and refolding simulation after downward pressure jump, constant temperature was maintained using Langevin dynamics with a damping constant of 1.0 ps^{-1}. Multiple time stepping was employed, with an integration time step of 2.0 fs, short-range forces evaluated every time step, and long-range electrostatics evaluated every three time steps. Cut-off for short-range non-bonded interactions was 8.0 Å, with shifting beginning at 7 Å; long-range electrostatics was calculated using the particle-mesh Ewald method[293]. All protein bonds involving hydrogen were constrained using RATTLE[294] and water geometries were maintained using SETTLE[295].

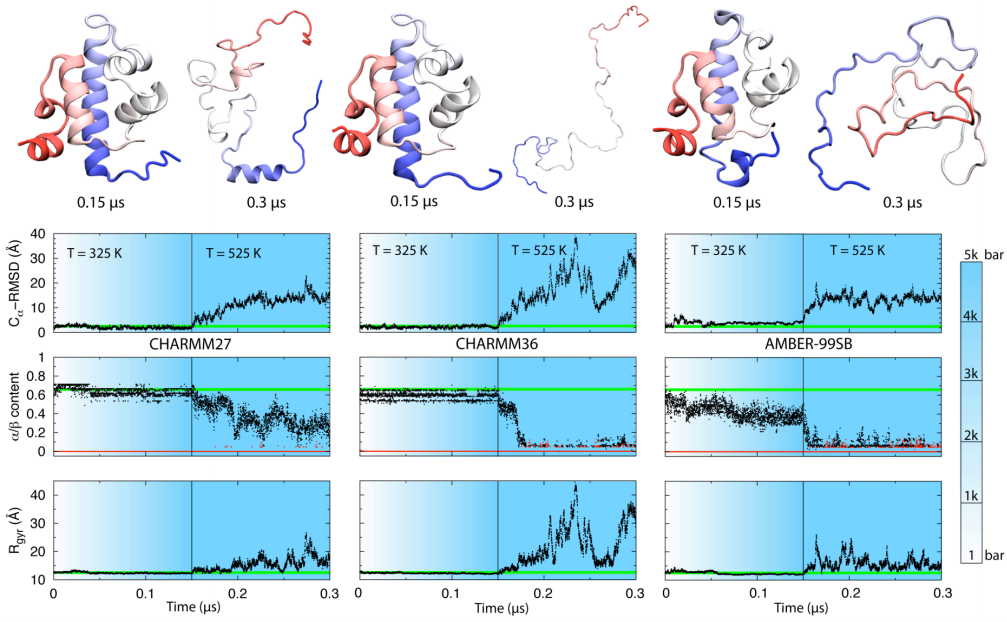


Figure C.8: Generation of λ-repressor P-T unfolded initial states using three different force fields: CHARMM27[35], CHARMM36[36], and AMBER299SB[37]. C_α-RMSD values have been calculated relative to the crystal structure 3KZ3[38]. α/β content is the fraction of residues that are in the α-helical conformation (black dot) and β-sheet conformation (red dot). R_{gyr} refers to the radius of gyration. The average native values, calculated from a 150-ns equilibrium simulation of the native structure at T = 325 K and P = 1 bar, are shown as green solid lines. The unfolding simulation started from the crystal structure λ-repressor fragment at T = 325 K and the pressure gradually increased in 0.15 μs from 1 bar to 5 kbar as indicated by the colored background; then the temperature was increased to 525 K and maintained for another 0.15 μs to completely unfold the protein while keeping the pressure at 5 kbar. The resulting λ-repressor conformation at the end of the pressurization and heating steps are shown at the top. Charmm27 has residual helical structure, Charmm36 and Amber299SB do not.

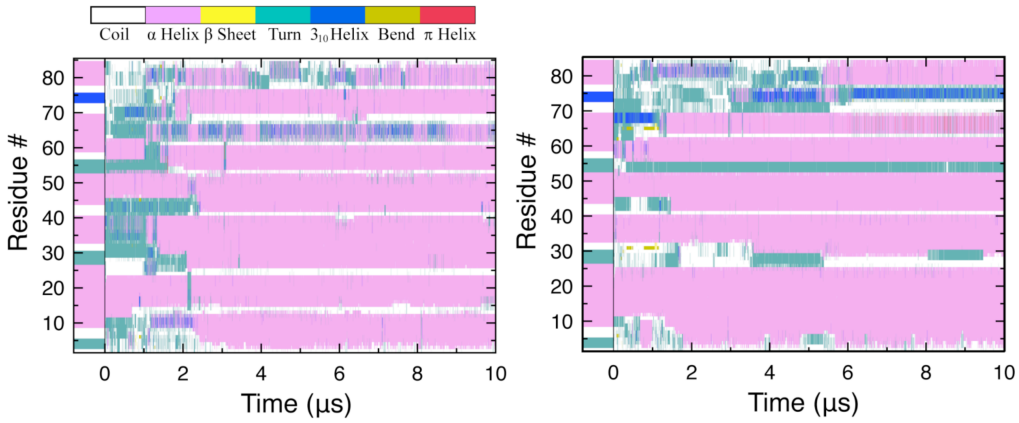


Figure C.9: Secondary structure throughout the two refolding trajectories using CHARMM27 force eld[35]. The secondary structure of the crystal structure is shown at the beginning as a reference. Color coding for secondary structure is shown at the top.

C.5.3 Simulation on the Anton platform

The refolding simulations were carried out on the Anton platform[296]. Multiple time stepping was employed, with an integration time step of 2.0 fs, short-range forces evaluated every time step, and long-range electrostatics evaluated every three time steps. Short-range non-bonded interactions were cut off at 9.26 Å; long range electrostatics was calculated using the k-Gaussian Split Ewald method[293] with a 64 × 64 × 64 grid. All bonds involving hydrogen atoms were constrained using SHAKE. The performance on Anton for our system using the above parameters was 5.0 μs/day.

C.5.4 Additional discussion of CHARMM27 simulation results

Protein unfolding under high pressure is a slow process which can take minutes and even hours (the experiments take ∼1/2 hour to raise the pressure to 1.2 kbar). As expected, the λ*YA did not change much when the pressure increased from 1 bar to 5 kbar in 0.15 μs. The high pressure slightly disrupted the helical structure (Figure C.8). The protein became more compact and structurally closer to the crystallized native state at high pressure than what observed in the equilibrium simulation of the native state at T = 325 K and P = 1 bar (green line in Figure C.8). To accelerate the un-

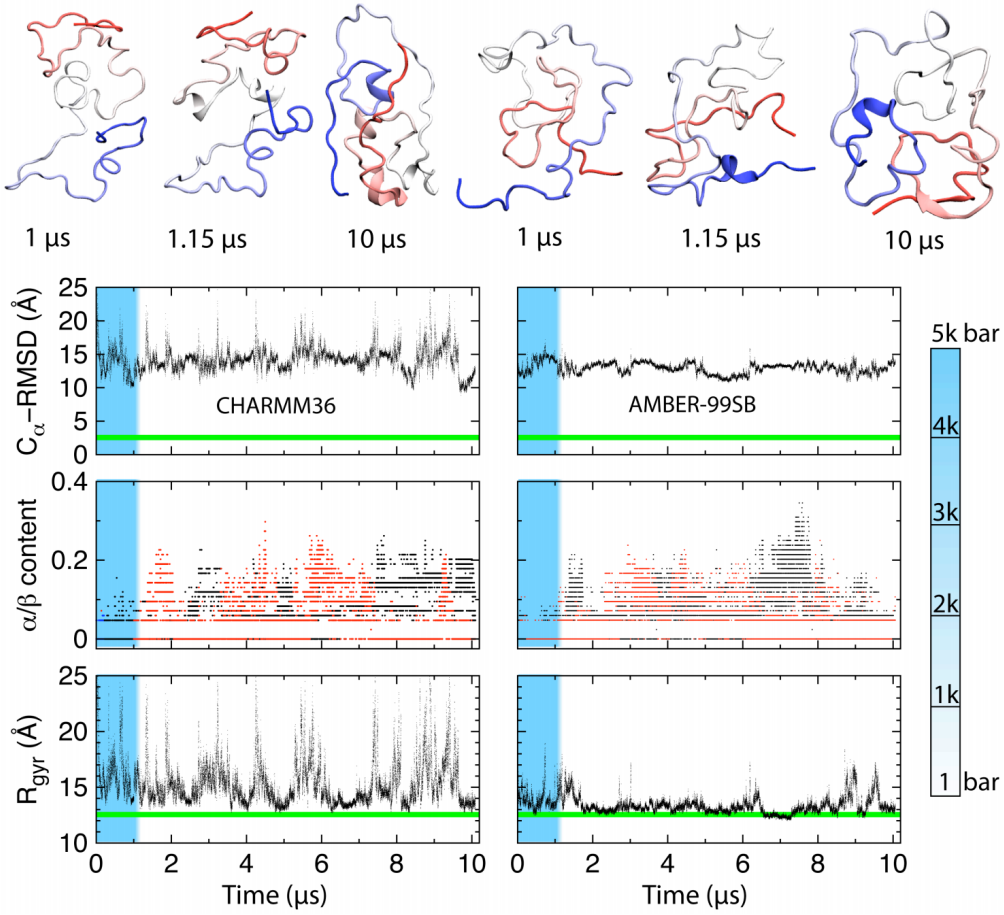


Figure C.10: Like Figure 4.5, simulation of λ*YA during pressure drop using CHARMM36[36] and AMBER299SB[37] force fields. The initial states for these simulations are the corresponding final states form Figure C.8. Top: structures from the two trajectories. The high pressure simulations start with 1 μs at 325 K and 5 kbar (blue zone), followed by 0.15 μs pressure drop to 1 bar (white zone). 8.85 μs of refolding were simulated at 1 bar and 325 K. C_α-root mean square deviations were calculated relative to the crystal structure 3KZ3[38]. α/β-content is the fraction of residues that are in the α-helical and β-sheet conformation, respectively. R_{gyr} is the unsolvated radius of gyration. The native values are dened by the mean values (green solid line) of the 150-ns equilibrium simulation of the native structure at T = 325 K and P = 1 bar. Note that the CHARMM36 structure (R_{gyr}) fluctuates more than any other we simulated.

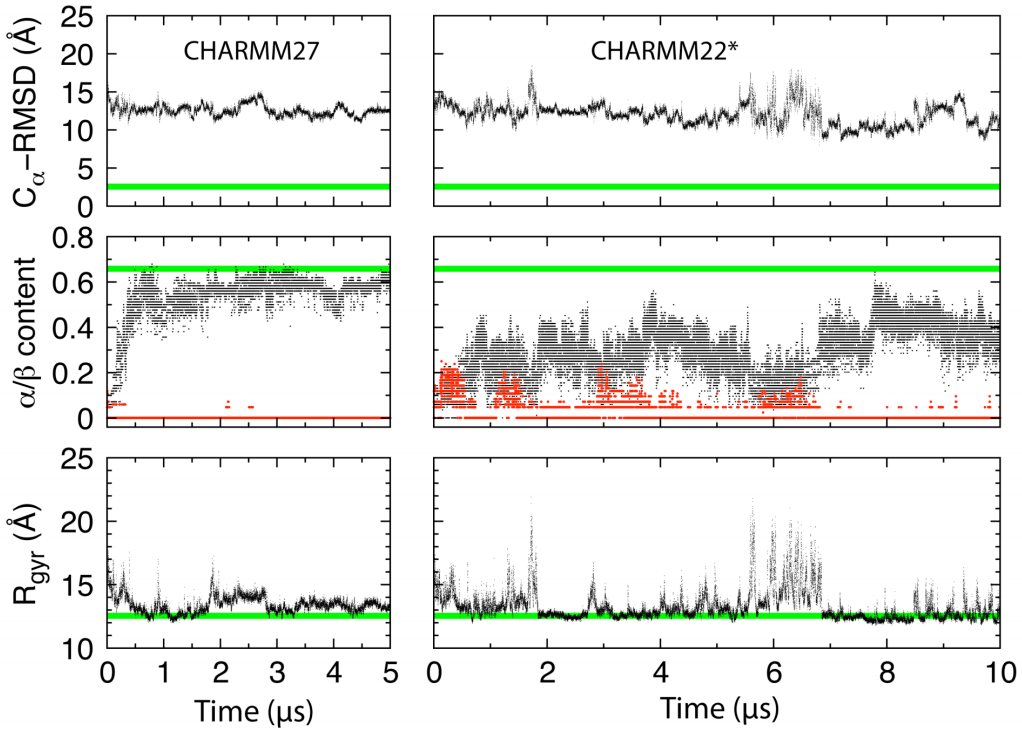


Figure C.11: Simulation of λ*YA refolding using CHARMM27[35] and CHARMM22*[39] force elds. Both simulations start from the AMBER299SB force eld in Figure C.3. C_α-root mean square deviations were calculated relative to the crystal structure 3KZ3[38]. α/β-content is the fraction of residues that are in the α-helical and β-sheet conformation, respectively. R_{gyr} is the radius of gyration. The native values are dened by the mean values (green solid line) of the 150-ns equilibrium simulation of the native structure at T = 325K and P = 1 bar.

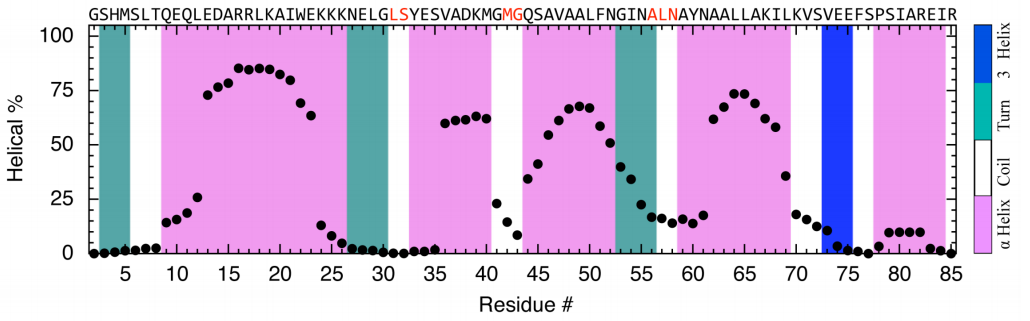


Figure C.12: Residue-specic α-helical propensity of the refolding simulation using CHARMM22*[39] in Figure C.11. The helical percentage was dened as the time percentage each residue spent in a α-helical conformation during the last 8 μs of refolding simulation. The secondary structure of the crystal structure is shown as a color-coded background, and the sequence at the top highlights turn/coil residues with >75% helix content in the simulations using CHARMM27 force eld (see Figure 4.6). Two of the loop regions identified in Charmm27 simulations (in red) also have helix propensity in the Charmm22* simulations.

folding process in simulations, we introduced high temperature of 525 K into the system. The protein completely unfolded in another 0.15 μs with a maximum root-mean-square deviation (RMSD) of 23 Å. The α-helical dropped to ∼20–30%, compared to ∼60–70% in the native state. A high-pressure and high-temperature denatured state at the end of the unfolding simulation is shown in Figure C.8. The high-pressure-high-temperature denatured state at the end of the unfolding simulation was taken as starting structure for two independent high pressure equilibrium simulations (1 μs each). The pressure then dropped from 5 kbar to 1 bar in 0.15 μs. The resulting high pressure denatured states from such simulations can have more α-helical element as shown in the second set of simulation. Although neither simulations refolding into the native state, the two simulations exhibited two types of signature events in the early stage of protein folding: in the rst set of simulation, the protein collapsed into more compact structures as evidenced by the decrease of radius of gyration R_{gyr}; in the second set of simulation, the protein undergoes large structure transitions as suggested by the spikes in the time course of the R_{gyr}.

APPENDIX D

SUPPLEMENTARY MATERIAL FOR "MULTIPROBE MAPPING REVEALS THE FAST FOLDING MECHANISM OF λ-REPRESSOR FRAGMENT 6–85"

Table D.1: Fitting parameters for the equilibrium temperature melt probed by fluorescence.

		Parameters	
	Mutants	T_m (°C)	$\partial\Delta G/\partial T$ (J mol^{-1} K^{-1})
Mean wavelength	λ_{12}	69 ± 1	132 ± 2
	λ_{13}	53 ± 1	57 ± 2
	λ_{32}	49 ± 1	59 ± 4
	λ_{42}	54 ± 4	27 ± 4
SVD trend #1	λ_{12}	–	–
	λ_{13}	50 ± 1	152 ± 29
	λ_{32}	47 ± 1	144 ± 33
	λ_{42}	36 ± 3	33 ± 3
SVD trend #2	λ_{12}	–	–
	λ_{13}	55 ± 1	98 ± 2
	λ_{32}	53 ± 1	108 ± 5
	λ_{42}	37 ± 2	40 ± 2

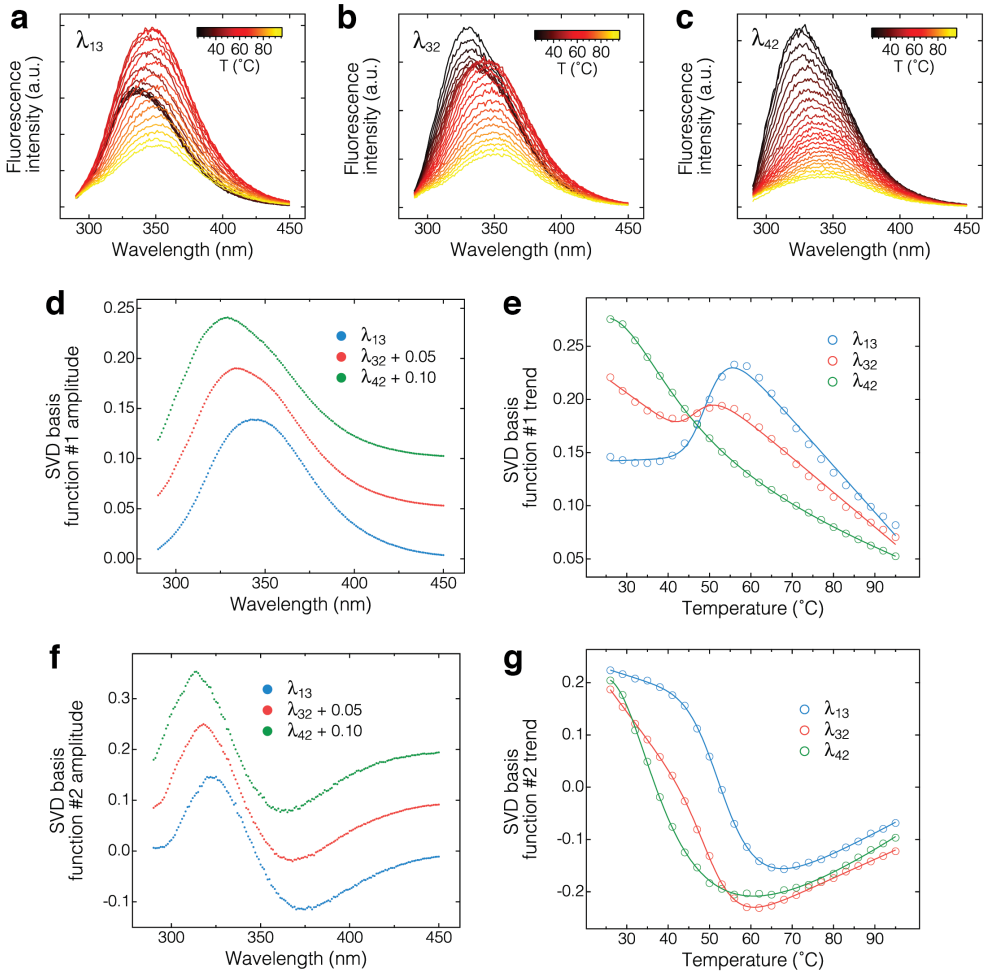


Figure D.1: Temperature melts of the new lambda repressor mutants probed by fluorescence spectroscopy. (a–c) Raw data for λ_{13}, λ_{32}, and λ_{42}, respectively. The spectra were taken from 23 °C to 95 °C in 3 °C increments. (d) Singular value decomposition basis functions 1. (e) Singular value decomposition basis function 1 trends. Note that both λ_{13} and λ_{32} show an increase in fluorescence intensity upon unfolding, while λ_{42} does not. (d) Singular value decomposition basis functions 2. (e) Singular value decomposition basis function 2 trends.

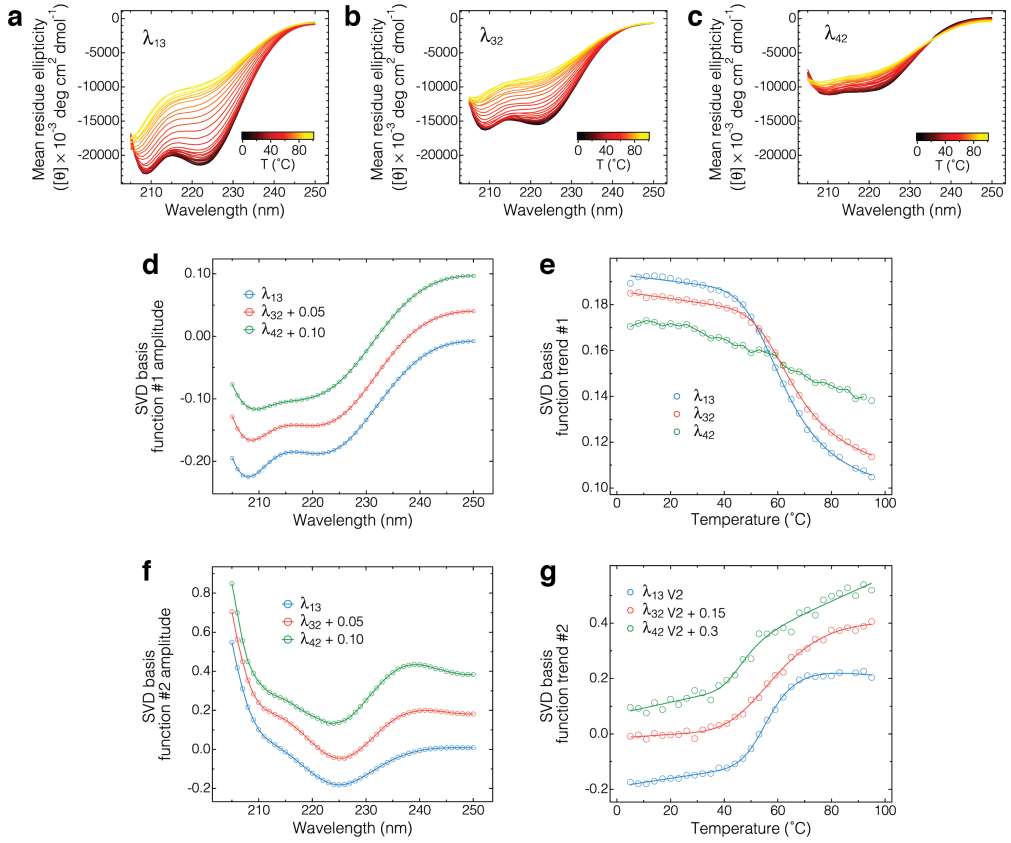


Figure D.2: Temperature melts of the new lambda repressor mutants probed by circular dichroism spectroscopy. (a–c) Raw data for λ_{13}, λ_{32}, and λ_{42}, respectively. The spectra were taken from 5 °C to 95 °C in 3 °C increments. (d) Singular value decomposition basis functions 1. (e) Singular value decomposition basis function 1 trends. Note that both λ_{13} and λ_{32} show a cooperative transition upon unfolding, while λ_{42} does not. (d) Singular value decomposition basis functions 2. (e) Singular value decomposition basis function 2 trends.

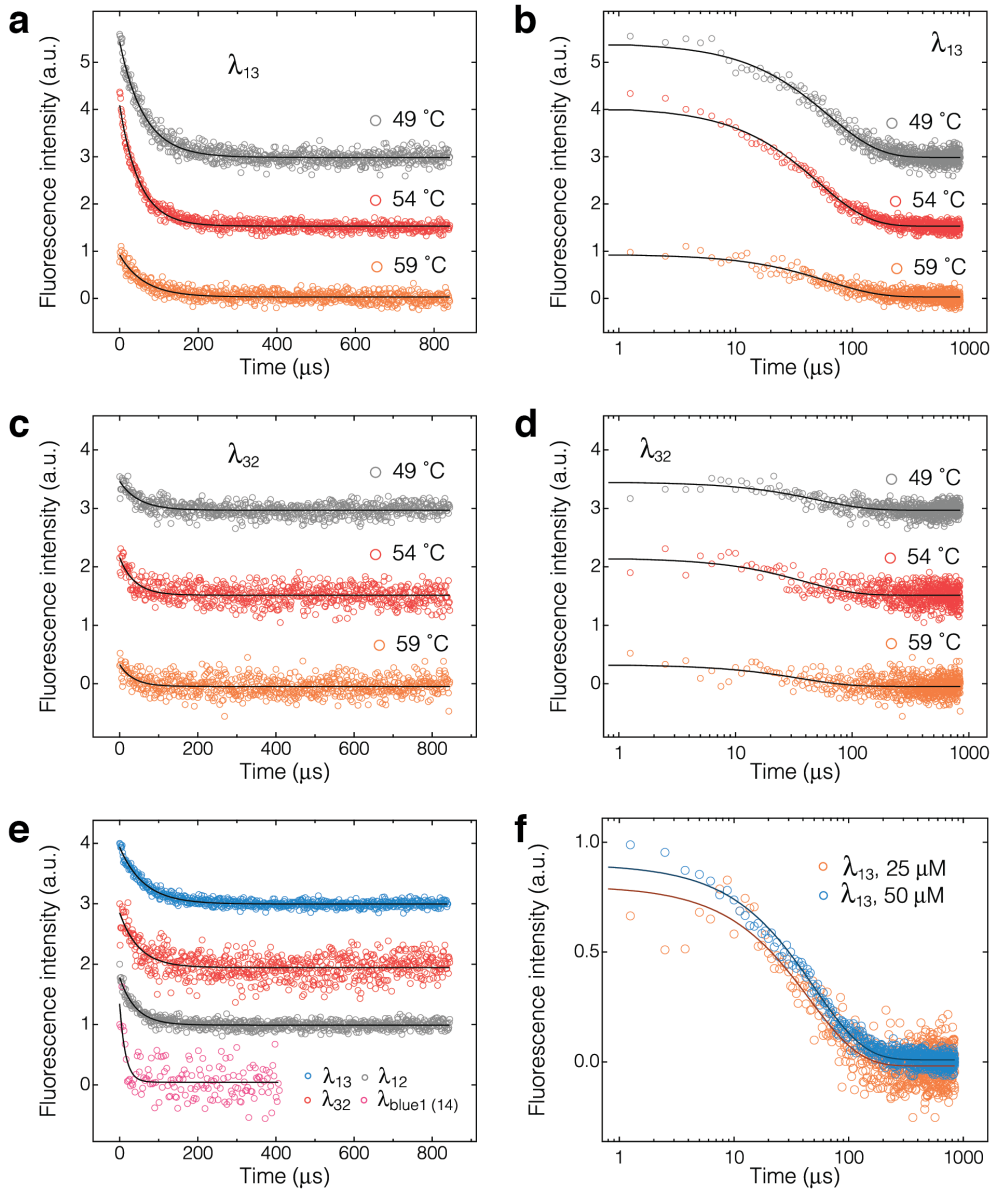


Figure D.3: Kinetic traces of lambda repressor mutants observed after a 9 °C temperature jump. (a) Kinetic traces of $lambda_{13}$ at various temperatures. (b) Same data as in (a) plotted against logarithmic time axis. (c) Kinetic traces of $lambda_{32}$ at various temperatures. (d) Same data as in (c) plotted against logarithmic time axis. (e) Same data as in the kinetics figure in Chapter 4 but plotted against the linear timescale. (f) Concentration dependence of $lambda_{13}$ kinetics. A slight slowdown is observed when the concentration is doubled from 25 μM to 50 μM.

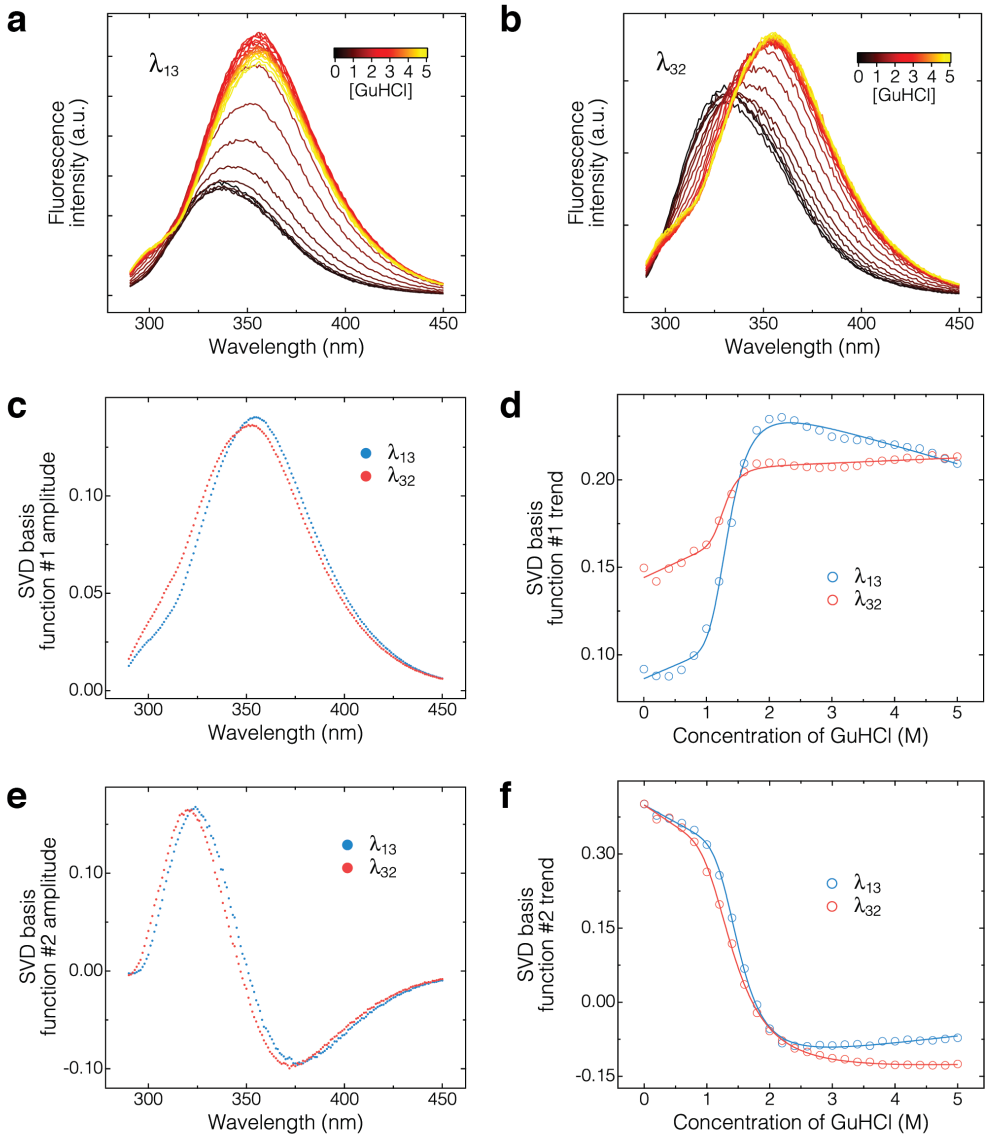


Figure D.4: Guanidine hydrochloride titrations of λ_{13} and λ_{32} probed by fluorescence spectroscopy. (a, b) Raw data for λ_{13} and λ_{32}, respectively. The spectra were taken from 0 M GuHCl to 5 M GuHCl in 0.2 M GuHCl increments. (c) Singular value decomposition basis functions 1 for both data sets. (d) Singular value decomposition basis function 1 trends. (e) Singular value decomposition basis functions 2 for both data sets. (f) Singular value decomposition basis function 2 trends.

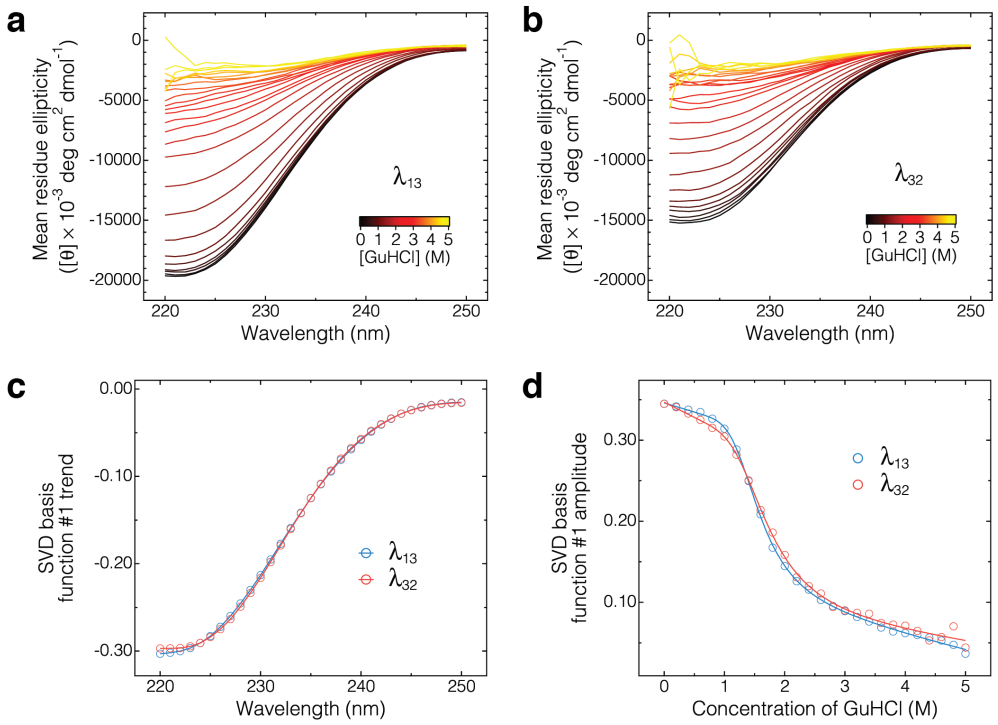


Figure D.5: Guanidine hydrochloride titrations of λ_{13} and λ_{32} probed by circular dichroism spectroscopy. (a, b) Raw data for λ_{13} and λ_{32}, respectively. The spectra were taken from 0 M GuHCl to 5 M GuHCl in 0.2 M GuHCl increments. (c) Singular value decomposition basis functions 1 for both data sets. (d) Singular value decomposition basis function 1 trends.

Table D.2: Fitting parameters for the equilibrium temperature melt probed by circular dichroism

		Parameters	
	Mutants	T_m (°C)	$\partial\Delta G/\partial T$ (J mol^{-1} K^{-1})
Signal at 222 nm	λ_{12}	–	–
	λ_{13}	62 ± 1	63 ± 5
	λ_{32}	64 ± 2	65 ± 8
	λ_{42}	–	–
SVD trend #1	λ_{12}	–	–
	λ_{13}	63 ± 2	59 ± 6
	λ_{32}	65 ± 1	68 ± 5
	λ_{42}	–	–
SVD trend #2	λ_{12}	–	–
	λ_{13}	58 ± 1	84 ± 7
	λ_{32}	67 ± 8	39 ± 11
	λ_{42}	47 ± 2	90 ± 42

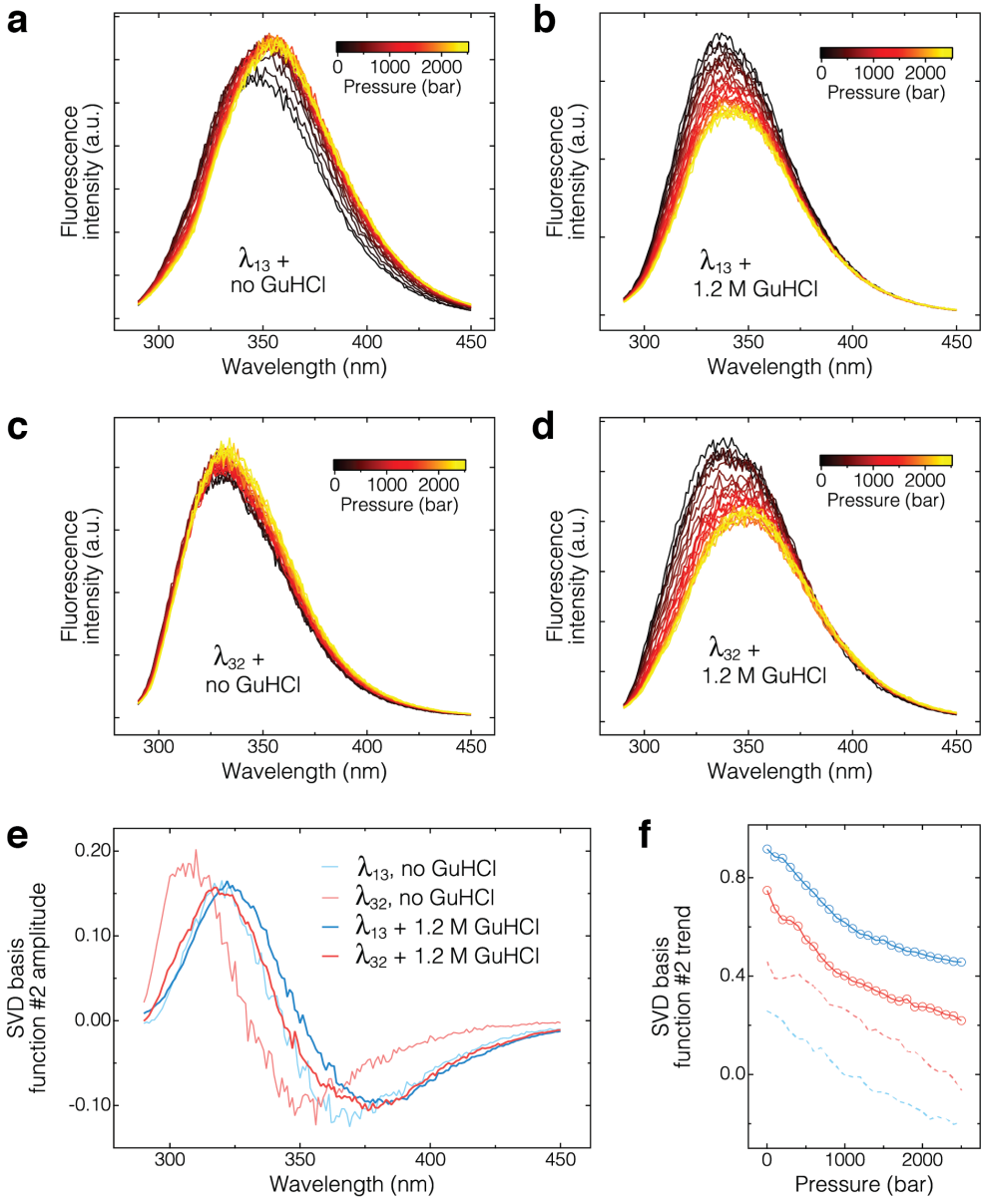


Figure D.6: Pressure melts of λ_{13} and λ_{32} with and without GuHCl probed by fluorescence spectroscopy. (a–d) Raw data for λ_{13} (a, b) and λ_{32} (c, d) with (b,d) and without (a, c) GuHCl. The spectra were taken from 1 bar to 2500 bar in 100 bar increments. (e) Singular value decomposition basis functions 2 for all four data sets. (f) Singular value decomposition basis function 2 trends.

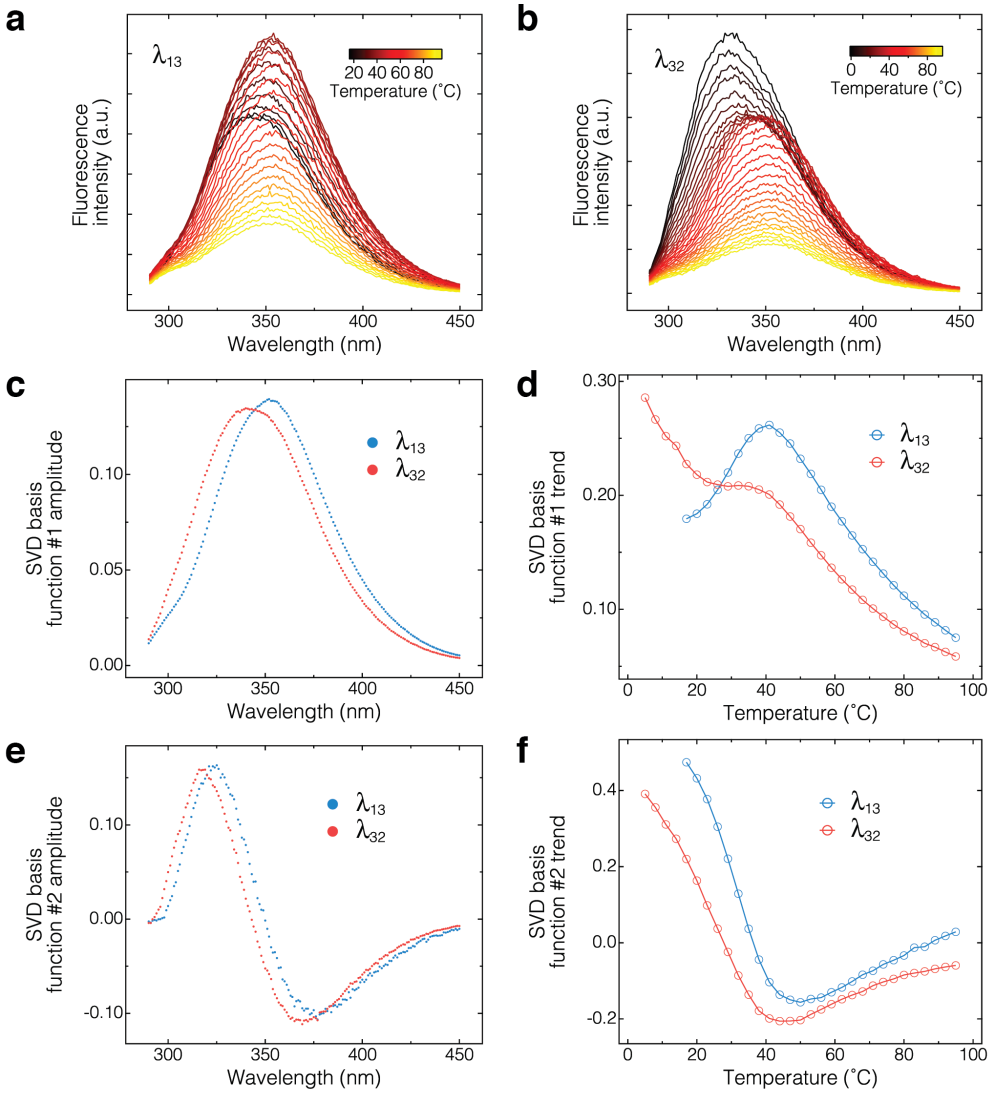


Figure D.7: Temperature melts of λ_{13} and λ_{32} in 1.2 m GuHCl probed by fluorescence spectroscopy. (a, b) Raw data for λ_{13} and λ_{32}, respectively. The spectra were taken in 3 °C increments. (c) Singular value decomposition basis functions 1. (d) Singular value decomposition basis function 1 trends. (e) Singular value decomposition basis functions 2. (f) Singular value decomposition basis function 2 trends.

Table D.3: Fitting parameters for the temperature jumps probed by tryptophan fluorescence lifetime. T^* is the temperature to which the temperature jump was done.

Mutants	T^* (°C)	τ_{obs}
λ_{13}	49 ± 1	62 ± 1
	54 ± 1	50.4 ± 0.8
	59 ± 1	62 ± 3
λ_{32}	49 ± 1	49 ± 5
	54 ± 1	39 ± 5
	59 ± 1	36 ± 7

Table D.4: Fitting parameters for the temperature jumps probed by tryptophan fluorescence lifetime. T^* is the temperature to which the temperature jump was done, which in each case is several degrees below T_{m} for each protein.

Mutants	T^* (°C)	τ_{obs}
$\lambda_{blue1}(14)$	63 ± 1	15 ± 4
λ_{12}	66 ± 1	43 ± 2
λ_{32}	49 ± 1	49 ± 5
λ_{13}	49 ± 1	62 ± 1

Table D.5: Fitting parameters for the temperature jumps probed by tryptophan fluorescence lifetime for λ_{13} at 54 °C.

Mutant	Concentration (μM)	τ_{obs} (μs)
λ_{13}	25	46 ± 2
λ_{13}	50	50.4 ± 0.8

Table D.6: Fitting parameters for the equilibrium temperature melt probed by circular dichroism

	Mutants	Parameters	
		c_{m} (°C)	$\partial\Delta G/\partial c$
Mean wavelength	λ_{12}	2.624 ± 0.003	10590 ± 82
	λ_{13}	1.149 ± 0.008	9581 ± 120
	λ_{32}	1.21 ± 0.01	7956 ± 148
Integrated intensity #1	λ_{12}	2.60 ± 0.02	11940 ± 898
	λ_{13}	1.33 ± 0.02	12330 ± 684
	λ_{32}	1.30 ± 0.04	17490 ± 3617
SVD trend #1	λ_{12}	–	–
	λ_{13}	1.33 ± 0.02	12400 ± 694
	λ_{32}	1.29 ± 0.04	18010 ± 3847
SVD trend #2	λ_{12}	–	–
	λ_{13}	1.50 ± 0.01	11090 ± 388
	λ_{32}	1.35 ± 0.03	8270 ± 422

Table D.7: Fitting parameters for the equilibrium temperature melt probed by fluorescence. Concentration of GuHCl is 1.2 M.

	Mutants	Parameters	
		T_{m} (°C)	$\partial\Delta G/\partial T$ (J mol^{-1} K^{-1})
Mean wavelength	λ_{13}	38 ± 1	24 ± 2
	λ_{32}	33 ± 1	39 ± 3

Table D.8: Fitting parameters for temperature and pressure jumps to the same final condition for λ_{13} and λ_{32}: [GuHCl] = 1.2 M, T = 23 °C. For λ_{12}, [GuHCl] = 2.4 M.

	τ_{obs} (μs)	
Mutants	T–jump	P–jump
λ_{12}	1851 ± 30	1439 ± 50
λ_{32}	536 ± 3	1618 ± 70
λ_{13}	1299 ± 11	–

APPENDIX E

SUPPLEMENTARY MATERIAL FOR "KINETIC AND THERMODYNAMICS EXPLORATION OF THE PRESSURE-TEMPERATURE PHASE DIAGRAM OF PHOSPHOGLYCERATE KINASE"

Table E.1: Fitting parameters for the equilibrium pressure melts probed by fluorescence and analyzed by mean wavelength at various temperatures.

Temperature (°C)	Parameters			
	P_{m1} (°C)	$\partial\Delta G_1/\partial P$	P_{m2} (°C)	$\partial\Delta G_2/\partial P$
9	674 ± 3	–38 ± 1	1193 ± 4	–14.2 ± 0.3
15	927 ± 7	–28 ± 1	1484 ± 12	–13.7 ± 0.9
23	1102 ± 11	–22 ± 2	1770 ± 58	–20 ± 8
30	1087 ± 4	–28 ± 1	–	–
36	905 ± 8	–29 ± 2	–	–

Table E.2: Fitting parameters for the equilibrium temperature melts probed by fluorescence and analyzed by mean wavelength at 1 and 500 bar.

Pressure (bar)	Parameters	
	T_{m1} (°C)	$\partial\Delta G_1/\partial T$
1	45.3 ± 0.1	281 ± 17
500	39.3 ± 0.2	308 ± 79

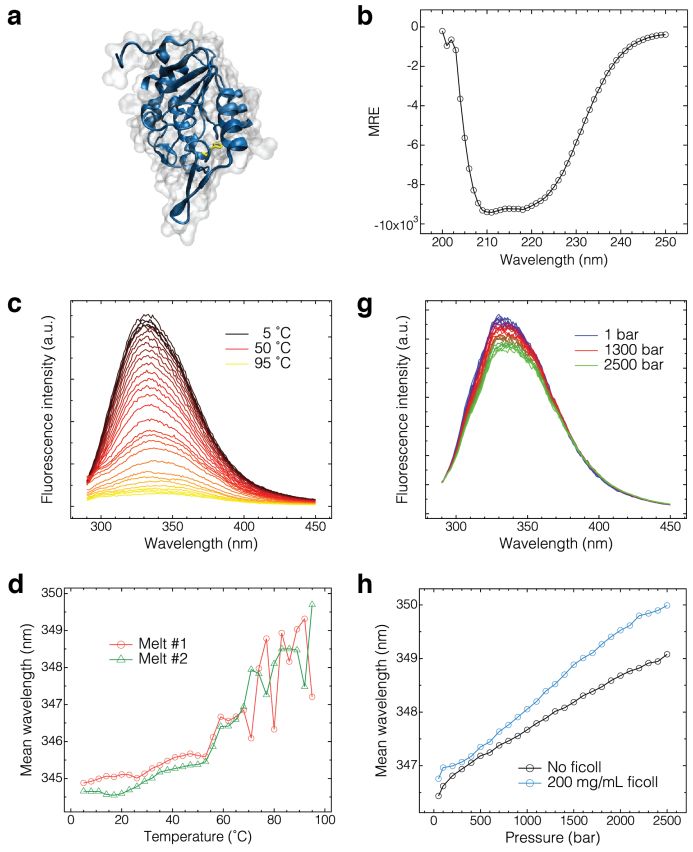


Figure E.1: Temperature- and pressure-induced thermodynamics of the N-terminal domain of PGK. (a) N-terminal domain of PGK. (b) Circular dichroism spectrum of the N-terminal domain of PGK showing substantial secondary structure content. (c, g) Fluorescence spectra of the N-terminal domain of PGK under temperature and pressure, respectively. (d, h) Spectral mean of spectra in (c) and (g), respectively.

Table E.3: Fitting parameters for the equilibrium pressure melts at 100 mg/mL ficoll probed by fluorescence and analyzed by mean wavelength at various temperatures.

Temperature (°C)	Parameters			
	P_{m1} (°C)	$\partial\Delta G_1/\partial P$	P_{m2} (°C)	$\partial\Delta G_2/\partial P$
9	801 ± 9	–55 ± 11	1336 ± 10	–13.5 ± 0.8
15	1044 ± 27	–37 ± 4	1524 ± 116	–16 ± 2
23	1204 ± 8	–24 ± 1	–	–
30	1185 ± 4	–35 ± 2	–	–
36	922 ± 3	–25.2 ± 0.6	–	–

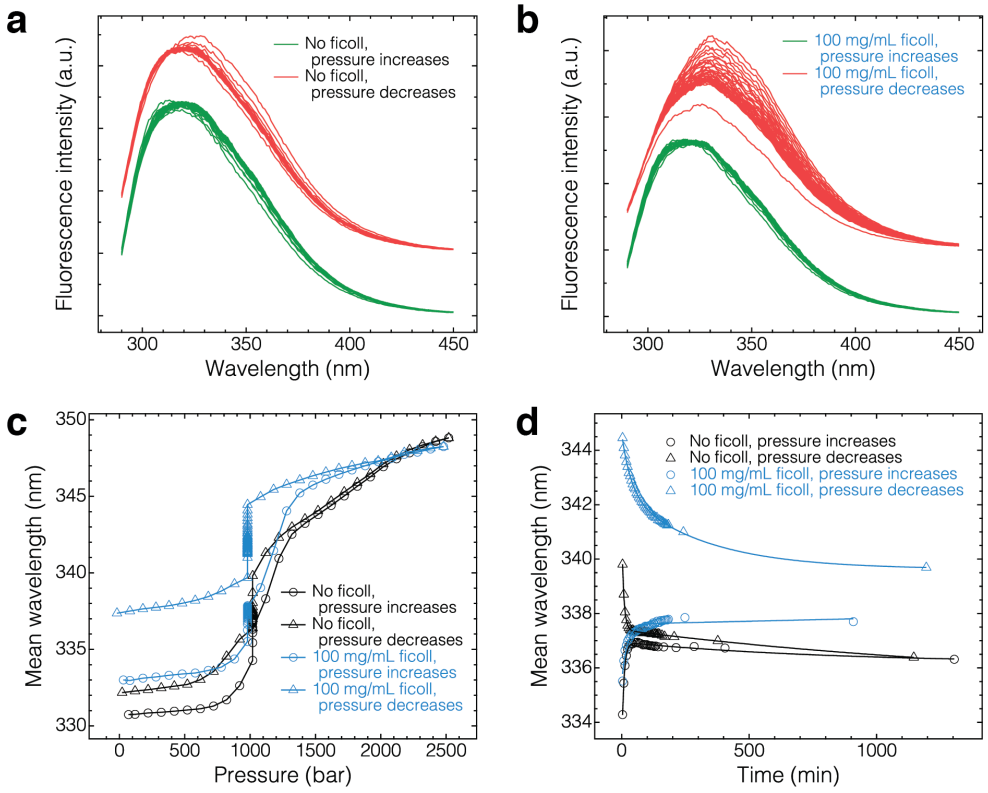


Figure E.2: Hysteresis in the pressure thermodynamics of PGK. (a) Change in fluorescence spectra on the way up and on the way down at 1000 bar without ficoll. (b) Change in fluorescence spectra on the way up and on the way down at 1000 bar in 100 mg/mL ficoll. (c) Hysteresis in the pressure thermodynamics of PGK with (blue) and without (black) ficoll on the way up (circles) and on the way down (triangles). (d) Kinetics at 1000 bar on the way up (circles) and on the way down (triangles) with (blue) and without (black) ficoll.

Table E.4: Fitting parameters for the equilibrium temperature melts at 100 mg/mL ficoll probed by fluorescence and analyzed by mean wavelength at 1 and 500 bar.

	Parameters	
Pressure (bar)	T_{m1} (°C)	$\partial\Delta G_1/\partial T$
1	46 ± 1	295 ± 26
500	39 ± 1	353 ± 97

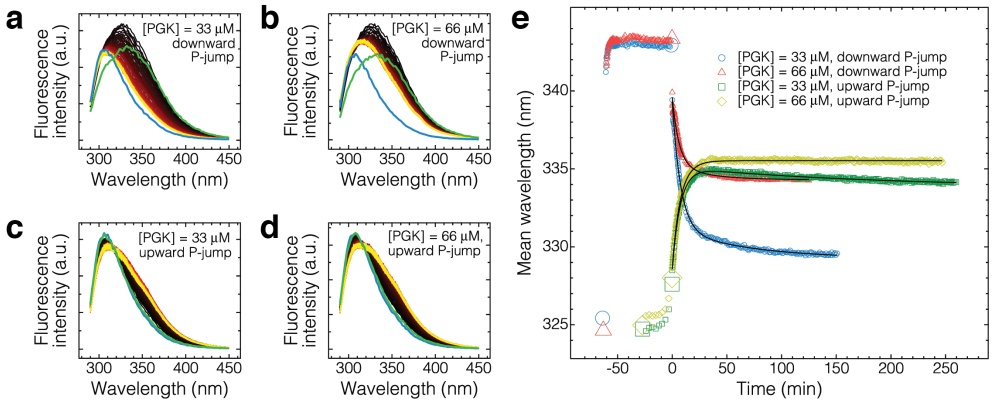


Figure E.3: Concentration dependence of pressure-induced kinetics of PGK. (a, b) Downward pressure jumps at 33 μM and 66 μM, respectively. Spectra in blue were taken at 1 bar before the pressure was increased. Spectra in green are the spectra taken right before the pressure was dropped. Spectra going from black to yellow represent a pressure-jump from 1600 bar to 1 bar. (c, d) Upward pressure jumps at 33 μM and 66 μM, respectively. Spectra in blue were taken at 1 bar before the pressure was increased. Spectra in green are the spectra taken right before the pressure was increased from 900 bar to 1000 bar. Spectra going from black to yellow represent a pressure-jump from 900 bar to 1000 bar. (e) Spectral mean representation of data in (a–d) showing all the data points when the pressure was increased before the pressure-jumps. Pressure-jumps occurred at $t = 0$. Solid lines are double-exponential fits.

Table E.5: Fitting parameters for the equilibrium pressure melts at 100 mg/mL ficoll probed by fluorescence and analyzed by mean wavelength at various temperatures.

[Ficoll] (mg/mL)	Parameters			
	P_{m1} (°C)	$\partial\Delta G_1/\partial P$	P_{m2} (°C)	$\partial\Delta G_2/\partial P$
0	1102 ± 11	–22 ± 2	1770 ± 58	–20 ± 8
50	1172 ± 29	–35 ± 12	–	–
100	1204 ± 8	–24 ± 1	–	–
150	1282 ± 4	–32 ± 1	–	–
200	1323 ± 11	–41 ± 6	–	–

APPENDIX F

SUPPLEMENTARY MATERIAL FOR "MECHANICAL MODELING AND COMPUTER SIMULATION OF PROTEIN FOLDING"

At the college freshman chemistry level this activity was piloted in one 3-hour session. The Manual included all the instructions for the activity and was posted as a pdf that the students accessed on their computers. The activity was punctured by mini-lectures 5-10 minutes in length at various times in order to clarify the terminology, relate the new material to previously covered topics regarding kinetics and thermodynamics, or to talk about potential real-world applications.

To conserve paper, the questions without instructions can be printed out for the students to fill out as they progress through the activity.

Typical student answers / suggested answers are given in blue in this document.

F.1 Materials

1. Two peptide models (folded and unfolded)
2. A ruler
3. A rectangular boundary made of Styrofoam
4. Computer program provided with this lesson

F.2 Mechanical model

F.2.1 Introduction and objective

We will study a 2-dimensional "protein" that is 10 amino acids long and is labeled with special tags at either end (Figure 1, the tags are not shown in the

figure but you can imagine them to be directly on top of amino acids 1 and 10). These tags allow researchers to measure the end-to-end distances of proteins, which are directly related to the conformation of a protein. Since the end-to-end distance is an important parameter for researchers, we will measure the end-to-end distances for the different conformations that our model proteins can attain. The next step will be to make sense of the collected data by distilling it to one or two meaningful statistical values to compare the data sets among students.

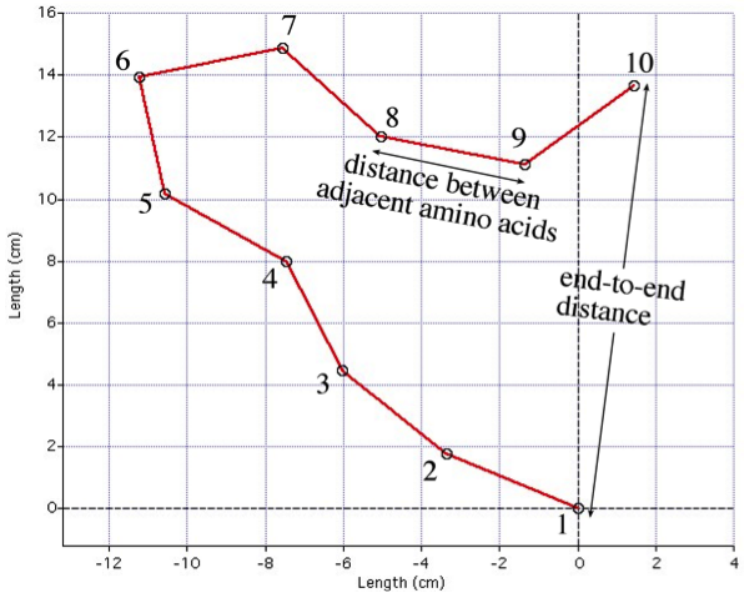


Figure F.1: A cartoon of a 10 amino acid model protein.

The first exercise is to measure the end-to-end distances of the model proteins after dropping them on a flat surface. This is a way of sampling the possible conformations that the model can achieve.

After you collect the data by dropping the model proteins and measuring the end-to-end distances of the resulting conformations, you will need to analyze your results. The idea is to present your data in such a way that you can retrieve maximum meaningful information out of it. In this particular case "meaningful" implies that you should be able to use that information to learn about the relationship between the studied model and the conformations that it samples and also to compare different models. Think about what information you might want to know about the conformational space that each model samples. First you would want to consider a distribution of end-to-end distances that you got. Of course, this distribution will not

be a perfect representation of the probability distribution function, which is a continuous distribution of occurrence probabilities, but it will give you an estimate. To visualize this distribution you will generate a histogram of your data. A histogram is a plot of the "number of events" as an ordinate (y-axis) and "end-to-end distance" as abscissa (x-axis). The data points on a histogram are usually shown as bars filling the space between the positive value on the y-axis (since of course there cannot be a negative number of events) and the x-axis. Each bar of the histogram has the same width called a "bin". A bin is a discrete interval of the x-axis within which all the observations are summed up and plotted. If you want to bin your data in steps of 1 cm, you would divide your data into groups, *e.g.* data points that fall between 0 cm and 1 cm, between 1 cm and 2 cm, between 2 cm and 3 cm, *etc.* You would then count how many of your measurements fall in each category and make a histogram out of it. Of course there are caveats like for example assigning a datum of exactly 1 cm to an interval 0–1 cm or 1–2 cm. In this case for rigorous analysis it must be specified explicitly how the intervals are defined. For example an interval can be defined as [n,n+1), i.e. including n but excluding n+1, in which case the 1 cm point will go into the 1–2 cm interval. You would also have to play with bin size before you get a histogram that is right for your purpose. On the one hand you dont want a bin size that is too large. An extreme case of that would be binning your entire data set into one point, which spans the entire range of end-to-end distances that your data covers. This binning would be useless because you know *a priori* that on the y-axis the value of that one bar will be equal to the number of trials that you did since all of the trials would be binned together. On the other hand you do not want to bin your data so finely that every datum is in its own bin because the sample size is quite small and the probability of two end-to-end distances being exactly 5.65 cm is extremely low. In that case if your bin size is, say, 0.01 cm, all data points will appear separately on the histogram and all bins will have 1 datum. That histogram is not telling you anything new compared to what your tabulated data was already saying. Thus, you are interested in finding an optimal binning size for your data. You can get a glimpse of the ways to decide on the best bin size from wikipeida (http://en.wikipedia.org/wiki/Histogram) under "Number of bins and width". We will just do it by trial-and-error here.

Once you got a histogram of your data you can start comparing a histogram

that you got for the "folded" *vs.* the "unfolded" protein model. At first the comparison can be visual. Does one histogram look like it is shifted with respect to the other? Are the shapes of the two histograms different? Perhaps one histogram is more "flat" with the events distributed more uniformly among the bins, while the other one is peaked sharply such that most of the events populated just one or two bins. These are qualitative observations that are good for initial comparison but to compare the two histograms quantitatively you would need to come up with one or two numbers that are easy to compare as opposed to comparing one data set with another or one histogram with another right off the bat. The two parameters that will help us in this comparison are the average and the full width at half the maximum (FWHM). An average can be obtained by taking the end-to-end distance that corresponds to every bin (*e.g.* you could decide on taking the center point, *i.e.* 1.5 for the 1–2 bin, and stay consistent with that choice throughout your analysis), multiplying it by the number of events in that bin, then adding up the obtained values for all the bins that have at least one data point in them, and dividing the result by the total number of events. The averages can then be compared quantitatively between the histograms by figuring out exactly by how much the average of one data set differs from that of the other. Of course, you can imagine two very different histograms that would have the same average value. Thus, an average is a good parameter to have but it is not sufficient by itself to define all the properties of a histogram. Another simple parameter that can help you is the FWHM. FWHM is a measure that works well for a singly peaked continuous distribution, *e.g.* a Gaussian distribution. To find FWHM you have to find the peak of the distribution by identifying the point that has the highest value on the y-axis, then divide that value by two and take the x-value difference of the points on the x-axis that correspond to the y-peak/2 value on the y-axis.

F.2.2 Activity 1: An unfolded protein

1. Take an unfolded protein model (i.e. the one in which all paper clips are flexible with respect to their neighbors), throw it upwards in the air and watch it fall on the desk.

2. Measure the end-to-end distance of the shape in which the model ended

up after Step 1 and record the end-to-end distances in Table F.1 below.

3. Repeat the first two steps 20 times.

Table F.1: End-to-end distances for the unfolded protein.

Trial number	End-to-end distance (cm)	Trial number	End-to-end distance (cm)
1.	17	11.	10
2.	12.5	12.	14
3.	6	13.	5
4.	18	14.	14
5.	13	15.	15
6.	5	16.	7
7.	10.5	17.	10
8.	8	18.	9
9.	7.5	19.	23
10.	10	20.	3.5

F.2.3 Activity 2: A folded protein

1. Do the same steps as in Section 2.2 but now use a folded protein model instead (*i.e.* the one where the first 5 amino acids are rigidly secured at 90° angles to imitate a partially folded conformation).

2. Use Table F.2 below to record your data.

Table F.2: End-to-end distances for the folded protein.

Trial number	End-to-end distance (cm)	Trial number	End-to-end distance (cm)
1.	15	11.	14.5
2.	16	12.	11
3.	11	13.	7
4.	10	14.	14
5.	9.5	15.	7.5
6.	20	16.	16.5
7.	25	17.	15
8.	8.5	18.	17
9.	15.5	19.	3.5
10.	13.5	20.	11.5

F.2.4 Data analysis

1. Use the data you collected to make two histograms of end-to-end distances for each model. You will end up with 4 histograms, two on each graph will correspond to the different model systems. It may be easier if you use different colors (*e.g.* pen and pencil or two colors of pen) for clarity. This way it will be easy for you to visually compare the histograms and get a qualitative idea of what the differences are. Use bin = 1 cm for the first histogram and bin = 5 cm for the second histogram.

   We need to compare the histograms that you generated in order to say what the differences are between the two data sets that we collected and hence the different protein structures. Perhaps you could simply look at the histograms and say that one is taller/wider than the other, or maybe one is overall shifted to the left/right with respect to the other. However it is hard to say exactly by *how much* one histogram differs from the other by mere visual inspection. For that we need to be able to get several key values out of our distributions and compare them instead:

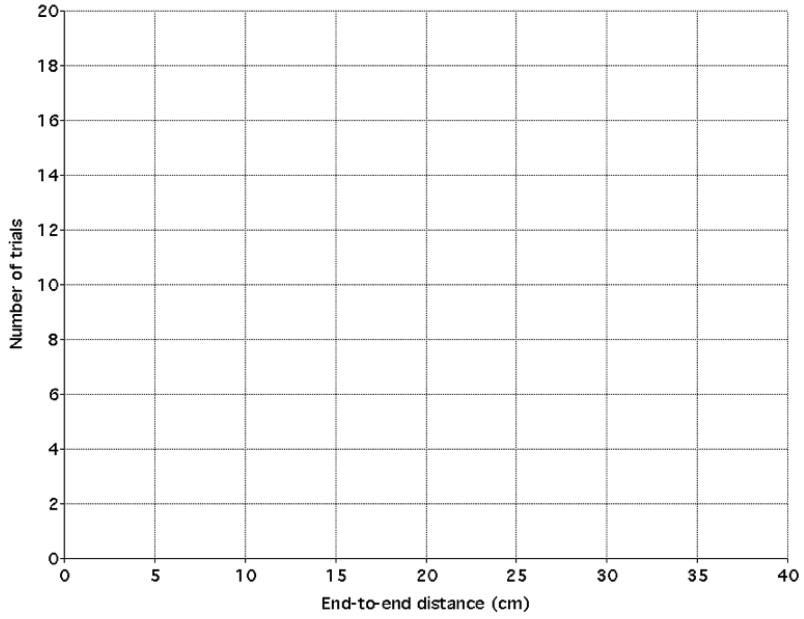


Figure F.2: Histogram 1. Bin = 1 cm.

2. For each histogram calculate the average. The average can be obtained

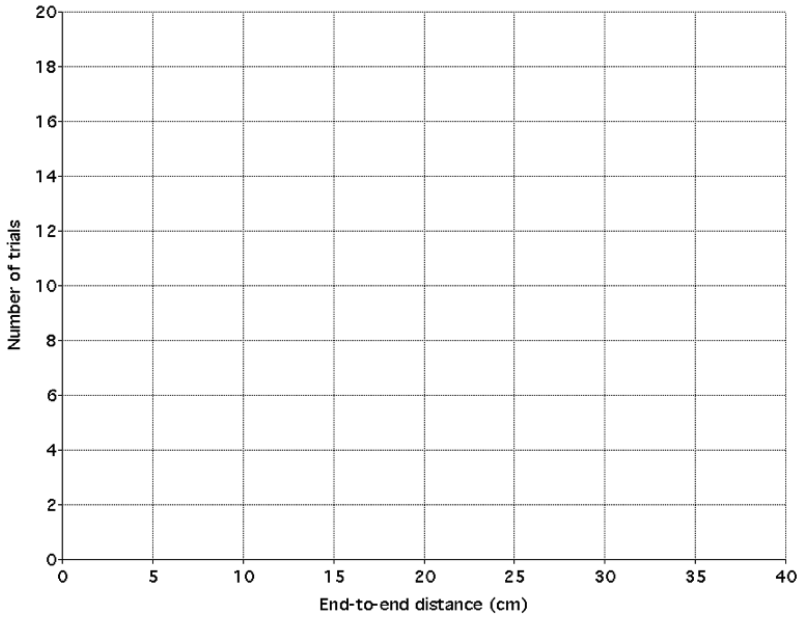


Figure F.3: Histogram 2. Bin = 5 cm.

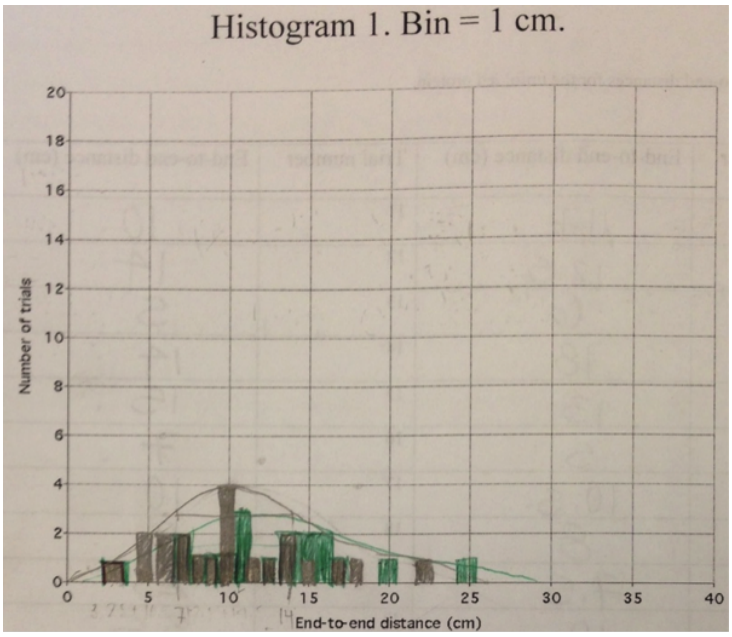


Figure F.4: Typical student data for Histogram 1.

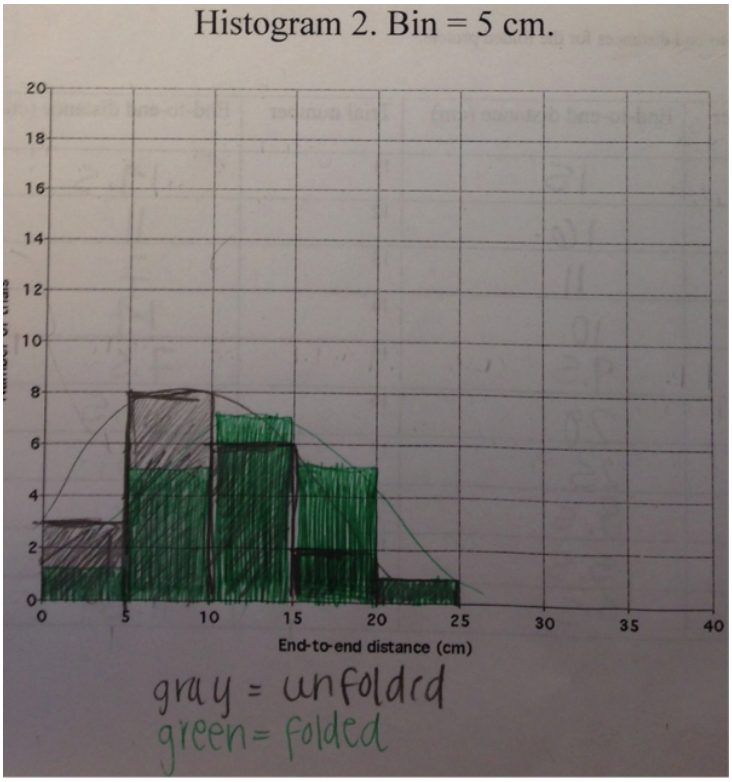


Figure F.5: Typical student data for Histogram 2.

by taking the end-to-end distance that corresponds to every bin (*e.g.* you could decide on taking the center point, *i.e.* 1.5 for the 1–2 bin, and stay consistent with that choice throughout your analysis), multiplying it by the number of events in that bin, then adding up the obtained values for all the bins that have at least one data point in them, and dividing the result by the total number of events.

Table F.3: End-to-end distances for the unfolded protein.

	1 cm bin	5 cm bin
Unfolded	11.9 cm	10 cm
Folded	12.3 cm	11.9 cm

3. You can imagine that two very different histograms could have the same average value. Thus, an average is a good parameter to have but it is not sufficient by itself to define all the properties of a histogram. Another simple parameter that can help you is the full width at half the maximum (FWHM). FWHM is a measure that works well for a singly peaked continuous distribution, *e.g.* a Gaussian distribution. To find

FWHM you have to find the peak of the distribution by identifying the point that has the highest value on the y-axis, then divide that value by two and take the x-value difference of the point on the x-axis that correspond to the y-peak/2 value on the y-axis. Calculate the FWHM for your histograms.

Table F.4: End-to-end distances for the unfolded protein.

	1 cm bin	5 cm bin
Unfolded	3.5 cm	5.5 cm
Folded	4.0 cm	6.3 cm

4. Now that you can compare the averages, what can you say about how the histograms are different and by how much?

   On both statistical measures, the unfolded protein ends are closer together than the folded ones by 1–2 cm. The 5 cm bins were less accurate than the 1 cm bins and the sample size was too small to be highly accurate.

5. How would you explain these differences based on the structures that your paper clip models were assuming when you were doing the measurements?

   The paper clip model for the folded proteins were rigid and forced a minimum distance for part of the protein, which caused the distances to be longer.

F.3 Computer simulation

Computer simulations are becoming increasingly more important in biological sciences as the processors become faster and cheaper and algorithms get better at approximating the behavior of biological systems. We will be using a simple computer program (Figure 6 is showing its graphical user interface) to simulate various structures that our paper clip protein can form and compare experimental results with a simulation. The program provides for a way of doing the same experiment as you already did in Activities 1 and 2 of Section 2, *i.e.* measuring end-to-end distances of a model protein and using them to build a histogram. However, while individually you could only do ∼50 trials in about half an hour, a computer can run through thousands of iterations in a matter of seconds and generate histograms that are much smoother than the ones you got. On the following pages you will find information on how to use the program.

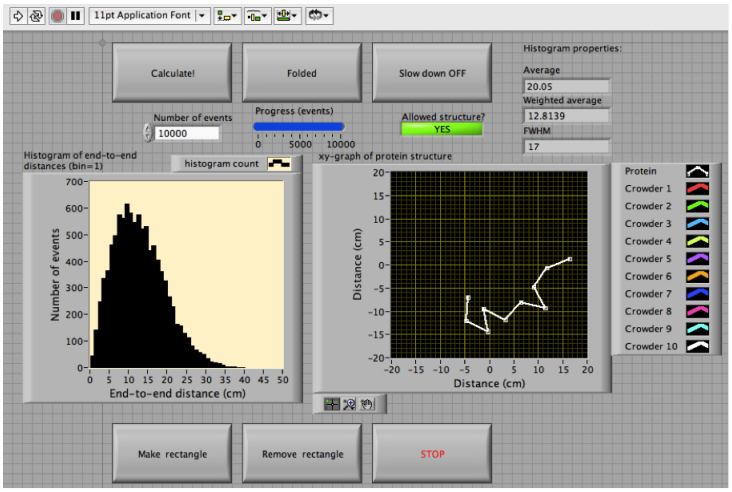
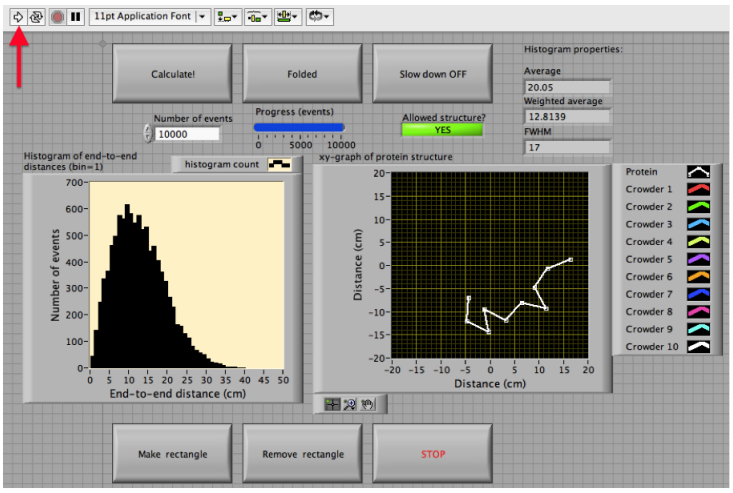


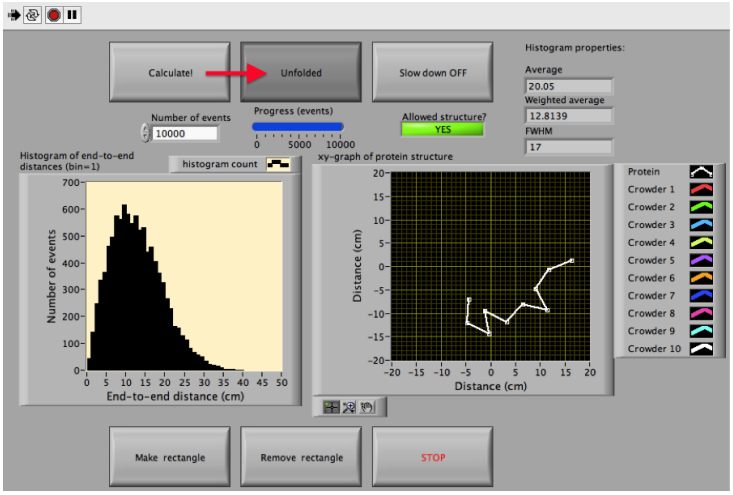
Figure F.6: Graphical user interface of the simulation program

F.3.1 Activity 1: An unfolded protein

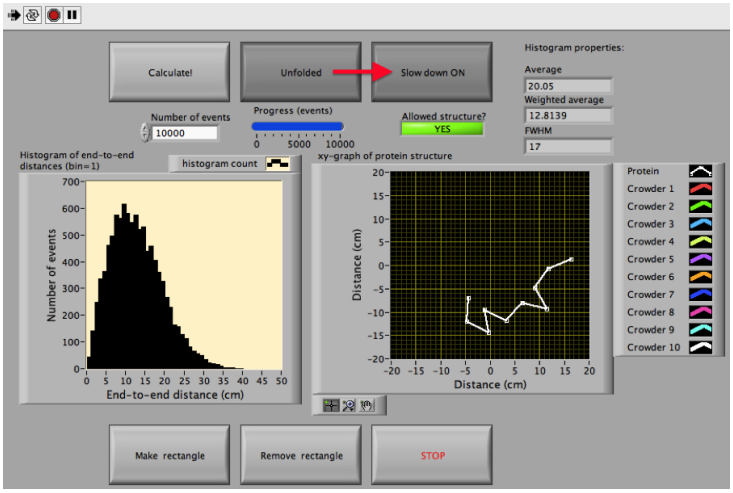
1. The program should already by loaded onto your computer. To start the program press the white arrow on the top left of the screen (highlighted with a red arrow on the figure below).



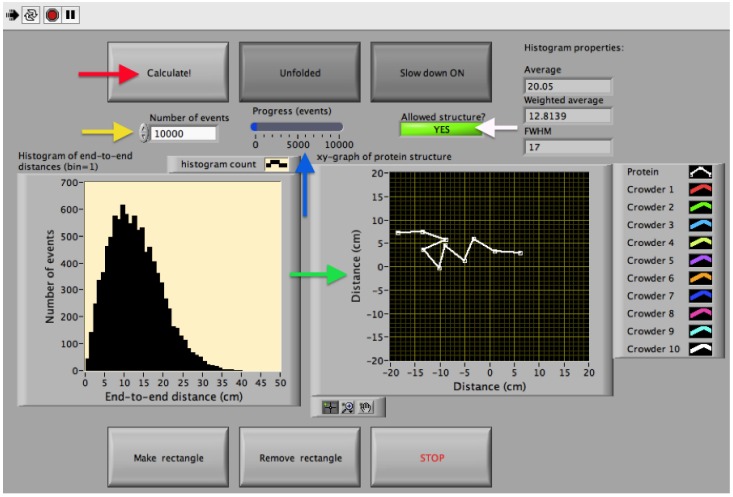
2. Once the program is running we can start looking into what kind of a histogram it will generate for an unfolded protein. First make sure that the middle button in the top row is pressed and the label on it reads "Unfolded" (red arrow). This means that the unfolded structure will be used in the analysis.



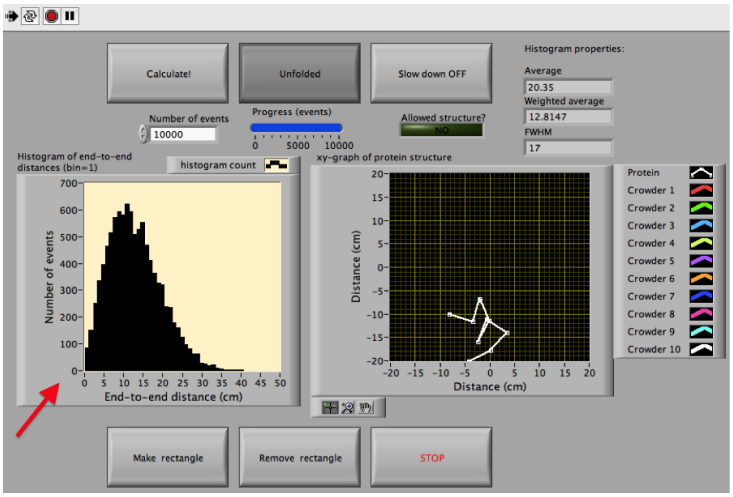
3. Now press the right-most button in the top row to switch it to the "Slow down ON" mode (red arrow). This will allow you to watch the structures change slowly to get an idea of what the program is doing.



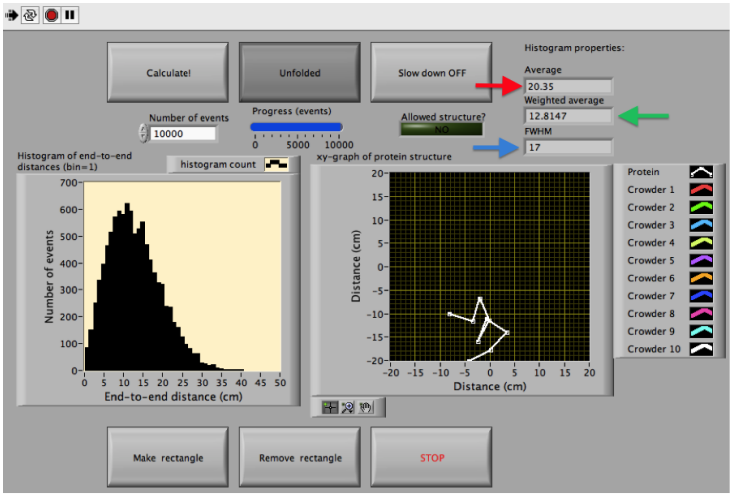
4. Now press the button that reads "Calculate!" – the left-most button in the top row (red arrow). You will see that protein structures on the "xy–graph of protein structure" in the middle of the screen (green arrow) will start changing as the program is going through the 10,000 events that it is required to complete ("Number of events" control is highlighted by the yellow arrow). You can follow the progress of the program using a "Progress (events)" indicator (blue arrow). Notice the "Allowed structure?" indicator (white arrow) is always green because there is no crowding in this case.



5. Once you have watched several structures change and you think you got a good look at them, press the "Slow down ON" button again to switch it to the "Slow down OFF" mode. The program will become much faster and will go through the rest of the structures very quickly. A histogram will also be generated (red arrow).



6. You will also find that some histogram properties have been calculated by the program: the average of all end-to-end distances, their weighted average, and a Full Width at Half the Maximum (FWHM). These values are highlighted with a red, a green, and a blue arrow, respectively. **Write these numbers down.**

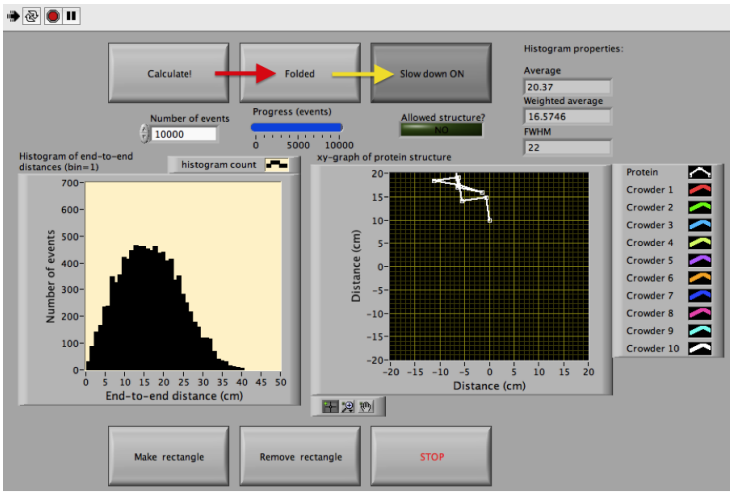


You can click "Calculate!" a couple of more times if you want, just to watch the program run.

Answers will vary based on simulation run. See screenshot above for example data.

F.3.2 Activity 2: A folded protein

1. Now click the "Unfolded" button to switch it to the "Folded" mode (red arrow) and click the "Slow down OFF" button to switch it to a "Slow down ON" mode (yellow arrow). Then click "Calculate!". The program is now running and if you look carefully you can see that the protein is now partially folded because the first five amino acids always come out to be at a 90° angle with respect to their neighboring amino acids.



2. Once you had a good look at the structures you can release the "Slow down ON" button so that the program becomes fast again and you can get your histogram and its properties for the folded structure. **Write these numbers down.**

   Answers will vary based on simulation run. See screenshot above for example data.

3. For the last exercise, just out of curiosity, let's play with the number of events for a little bit. Change the number of events from 10,000 to 100,000 and click "Calculate!". It will take the program a little longer to complete the simulation now that you requested ten times more structures to be sampled, but you can clearly see that the resulting histogram is very smooth and its properties are quite precise.

   You can also change the number of events to 20 – the same number that you actually went through in your homework with the paper clip models

– and run the simulation. You can see that the resulting histogram is not smooth, which means that its properties are not a very good representation of what the model protein is actually doing.

4. Stop the program by clicking the "Stop" button, the right-most button in the bottom row.

5. The main disadvantage of the models that we have been working with so far is that they can only populate one thermodynamic state: the folded protein model will always stay folded, no matter how you throw it, *ditto* the unfolded one. Thus, we can only get the information about how each state would behave individually but we have no idea what the interplay between the two states is like. Generally speaking real proteins do not stay in one state all the time, they switch from one state to another (*e.g.* from folded to unfolded, active to inactive, *etc.*) depending on the conditions, *i.e.* state variables, such as temperature, pressure, concentration, electrical potential, *etc.* To make our model more general you would have to change the paper clips (or come up with a totally new representation) such that the new model can be either folded or unfolded with a certain probability for being in each state. Describe a model that would have both the "folded" and the "unfolded' states integrated in it such that either one can show up. It doesnt matter if the probability is 50:50 or 20:80. What matters is that it is not 0:100.

   Answers will vary, but should include a way for the protein to switch between the folded and unfolded states.

F.4 Folding thermodynamics and kinetics

We have looked at a computer simulation that allowed us to get a better sampling of the end-to-end distances that the folded and the unfolded models were sampling by themselves and what the corresponding histograms would look like. This time around we need to do better because we already have several examples of models that can interconvert between the two states. If the end-to-end distances were measured for these new models, the final

histogram would likely be some linear combination of the histograms representative of each state. What we want now is a computer program that will do all the same things as the previous one did, except now we do not want to set the protein to be always folded or always unfolded but rather make it fold or unfold with predefined conversion rates determined by the folding rate constant, k_f, and the unfolding rate constant, k_u. The reaction we are studying is shown in Figure 7.

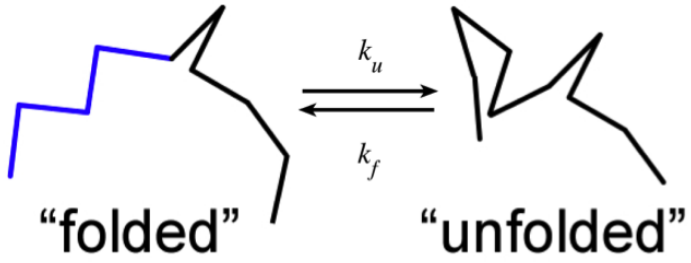


Figure F.7: Protein folding reaction of interest

A 1-dimensional free energy landcape for this reaction under folding bias, *i.e* $k_f > k_u$, looks something like the one in Figure 8.

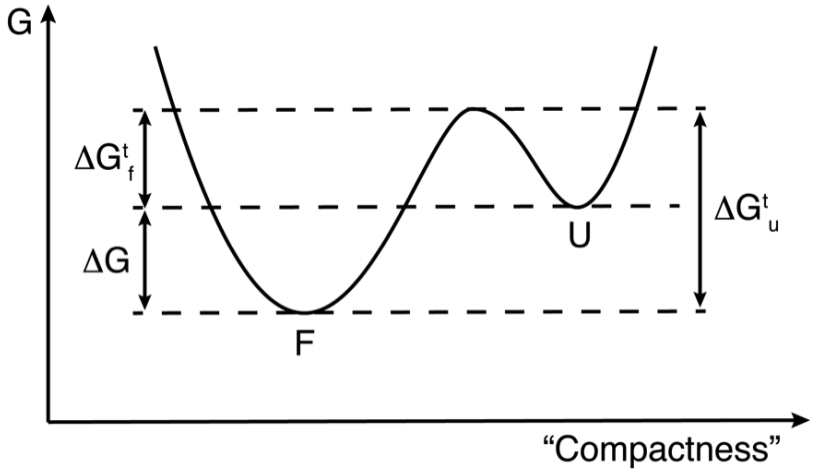


Figure F.8: Free energy landscape of the reaction of interest

In this case $k_f > k_u$ since $k_f = \frac{1}{\tau_f} = A \times e^{-\Delta G^\dagger_f / RT}$ according to the Arrhenius equation. Similarly, for unfolding $k_u = \frac{1}{\tau_u} = A \times e^{-\Delta G^\dagger_u / RT}$. In the figure $G^\dagger_u > G^\dagger_f$, so $k_f > k_u$.

In words, kinetically it is faster (the probability is higher) to convert from the unfolded state to the folded state because the barrier is smaller as com-

pared to the reverse reaction. Thermodynamically, folded state will be more populated at equilibrium because it is lower in free energy than the unfolded state.

$K_{eq} = \frac{[F]}{[U]} = \frac{k_f}{k_u} = e^{-\Delta G/RT}$, where $\Delta G = G_f - G_u < 0$,

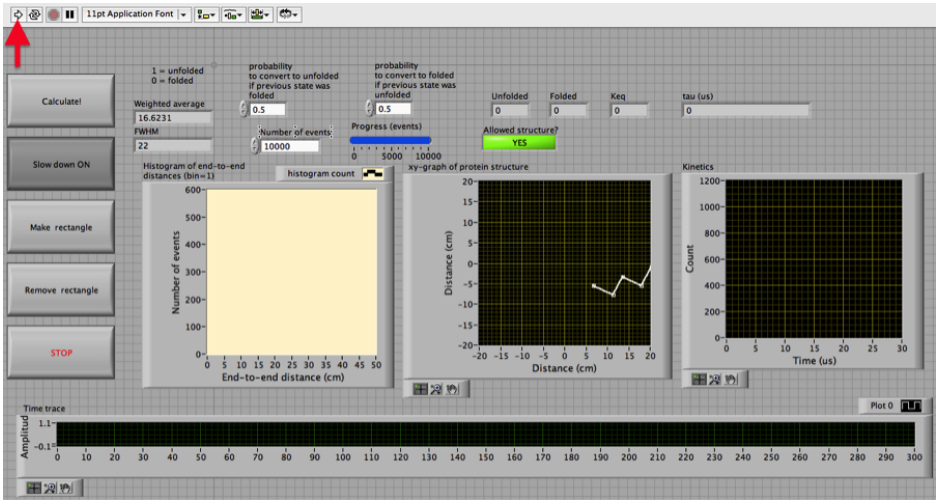
so $K_{eq} > 1$ and $[F] > [U]$.

Note the that the unfolded state looks like a shallow basin and is confined to a small region on the "compactness" reaction coordinate. What is the reason for this difference? Is your previous data in agreement with it?

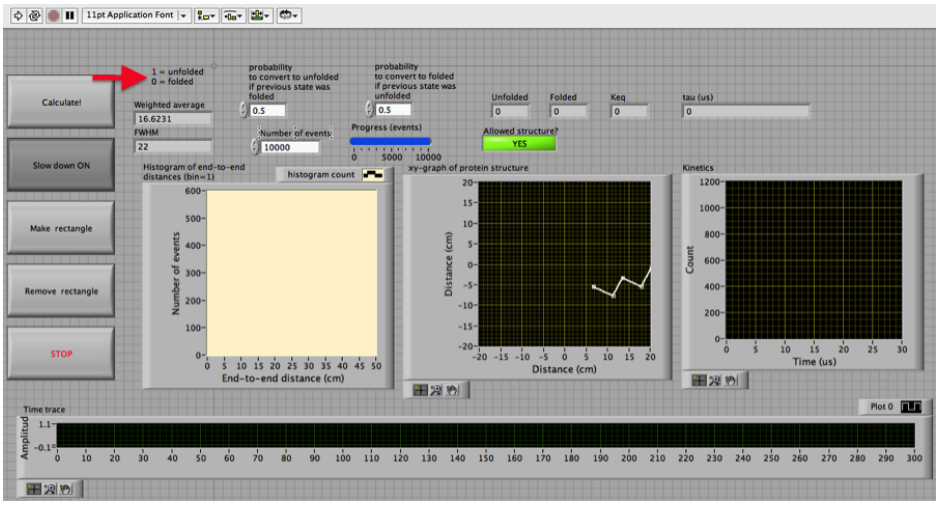
The model for the unfolded protein tended to land in a more compact state and had less variation in the end-to-end distance than the folded protein.

F.4.1 Activity 1: Computer simulation

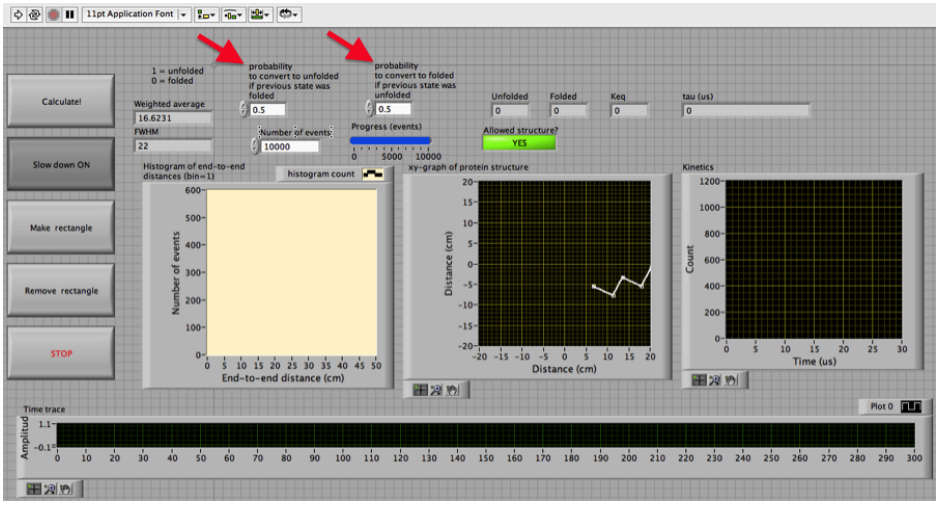
1. The program should already be loaded onto your computer. To start the program press the white arrow on the top left of the screen (highlighted with a red arrow on the figure below).



2. As you can see there are several new features in this program as compared to the previous one. At the top you can see that I defined the unfolded state as 1 and the folded state as 0 (highlighted with a red arrow on the figure below).

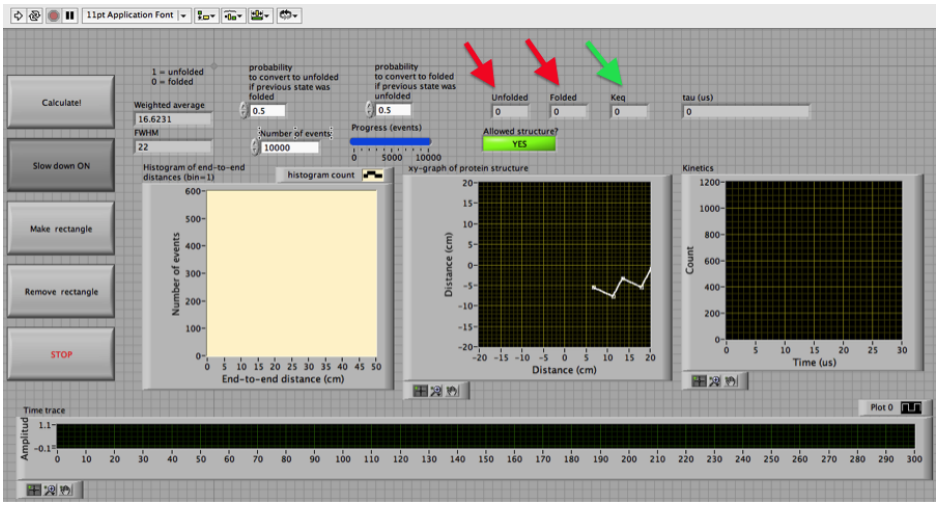


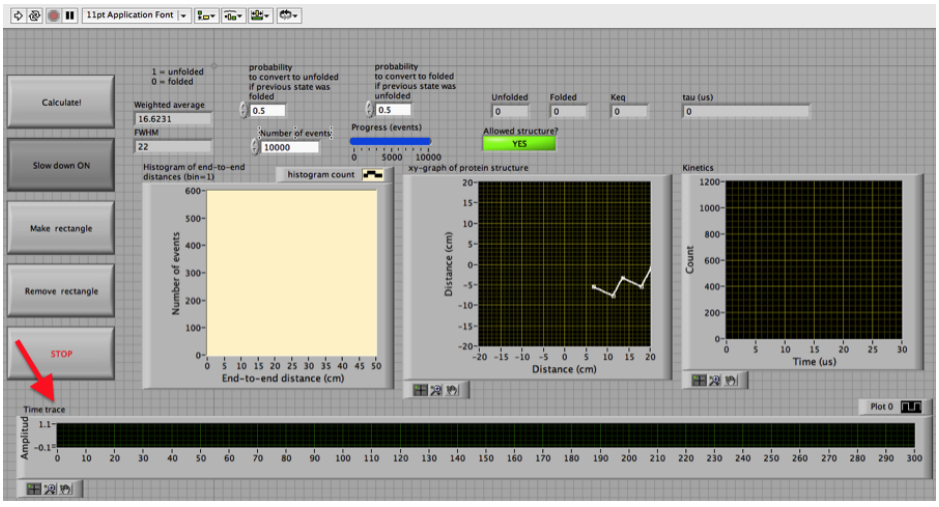
3. Further to the right you can see the probabilities for conversion from one state to the other based on which state the protein is in at the moment (highlighted with red arrows on the figure below). You can change those and we will play with them.



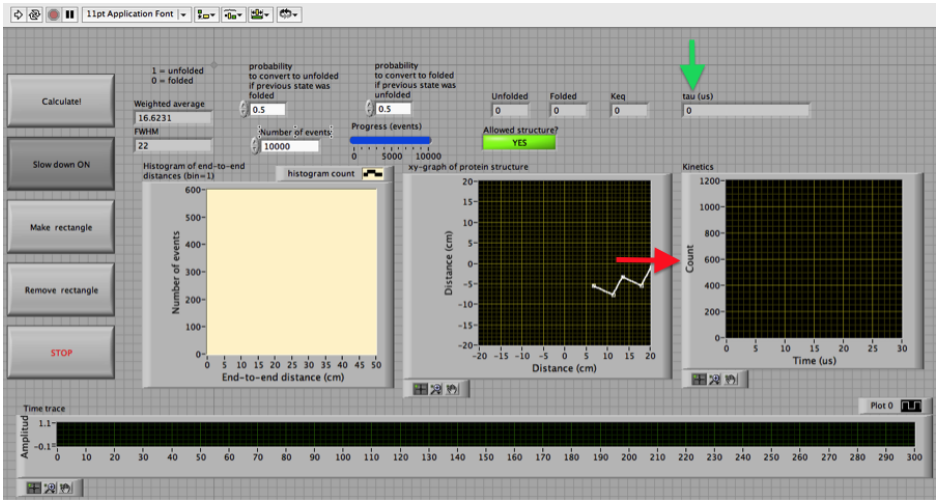
4. Even further to the right you can see the number indicators that will show you how many folded and how many unfolded conformations the program generated (highlighted with red arrows on the figure below). Remember that last time because all the conformations were identical you just had to set the number of iterations that you wanted the program to go through and you would be done. Now the total number of iterations (folded + unfolded) is determined by the user the same way but how this total number is distributed between the two states is determined by the conversion rates introduced above. From these two values it is easy to calculate $K_{eq} = \frac{[U]}{[F]}$ (highlighted with a green arrow on the figure below).

5. At the botom of the screen you can see a graph labeled "Time trace" (highlighted with a red arrow on the figure below). That graph is binary, *i.e.* it will show a value of 1 or 0, which will indicate whether a conformation is unfolded (1) or folded (0) as a function of "time". What I refer to as "time" here is actually just a conformation count but in order to talk about kinetics we have to talk about time. So you can imagine that our sampling rate is 1 MHz; so a new conformation appears every 1 μs.


11pt Application Font
Calculate!
Slow down ON
Make rectangle
Remove rectangle
STOP
1 = unfolded
0 = folded
probability to convert to unfolded if previous state was folded
probability to convert to folded if previous state was unfolded
Unfolded
Folded
Keq
tau (us)
Weighted average
16.6231
FWHM
22
Number of events:
10000
Progress (events)
Allowed structure?
YES
Histogram of end-to-end distances (bin=1)
histogram count
Number of events
End-to-end distance (cm)
xy-graph of protein structure
Distance (cm)
Kinetics
Count
Time (us)
Time trace
Amplitud
Plot 0

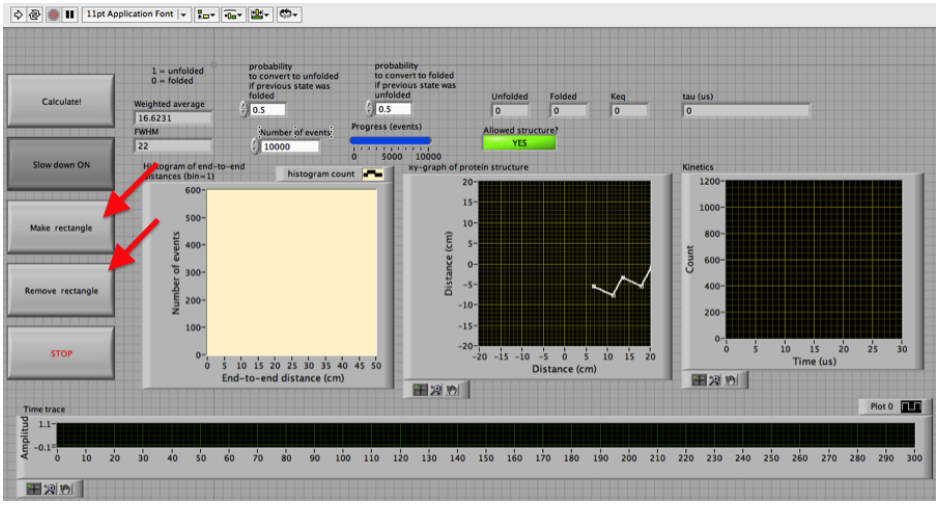

11pt Application Font
Calculate!
Slow down ON
Make rectangle
Remove rectangle
STOP
1 = unfolded
0 = folded
probability to convert to unfolded if previous state was folded
probability to convert to folded if previous state was unfolded
Unfolded
Folded
Keq
tau (us)
Weighted average
16.6231
FWHM
22
Number of events:
10000
Progress (events)
Allowed structure?
YES
Histogram of end-to-end distances (bin=1)
histogram count
Number of events
End-to-end distance (cm)
xy-graph of protein structure
Distance (cm)
Kinetics
Count
Time (us)
Time trace
Amplitud
Plot 0

6. The program keeps track of different conformations and in what order they show up on the screen. It then considers all the intervals for which the protein stayed folded and histograms them on the right (highlighted with a red arrow on the figure below). That histogram is then fitted with an exponential function Ce^{t/τ_u} and the time constant for the unfolding rate, τ_u, is displayed (highlighted with a green arrow on the figure below). A similar analysis can also be done for the folding rate by considering the intervals where the protein was unfolded. Which conversion probability out of the two that you can tweak do you think the time constant will be sensitive to?



The unfolding rate is $k_u = \frac{1}{\tau_u} = Ae^{-\Delta G_u^\dagger/RT}$ and the resulting graph is fitted to $Ce^{t/\tau_u} = Ce^{tk_u}$, so it is the barrier going from the folded state to the unfolded state (*i.e.* unfolding reaction), $\Delta G_u^\dagger$, that is important in this case. The time constant will be sensitive to "the probability to convert to the unfolded state if the previous state was folded".

7. There is another feature in this program. The two buttons dedicated to it are on the left panel; they a called "Make rectangle" and "Remove rectangle" (highlighted with a green arrow on the figure below). I will explain to you what it does later on.



8. Let's start playing with the program. Make sure that the "Slow down" button is ON and that both probabilities for folding and unfolding are 0.5. Now press "Calculate!". Protein conformations are going to start cycling on the screen and the time trace at the bottom will show a "snake" that corresponds to the jumping of a protein from one state to another. You can actually think of it as one protein changing conformations in time as opposed to a new protein showing up every iteration. If you wait long enough you will also see that the bars on the histogram start accumulating events and the histogram starts growing taller.

9. Now release the "Slow down" button. The program will dash through the rest of the conformations and build a histogram for you. Let's examine the equilibrium parameters that the program calculated. If you did everything correctly up to this point you must have gotten $K_{eq} \approx 1$. Why is the number not exactly equal to unity? What does $K_{eq} \approx 1$ imply in terms of the shape of the free energy landscape compared to the one shown in Figure 8? How could you tune the probabilities to get to the free energy landscape given in Figure 8?

   If $K_{eq} = 1$, then ΔG for the landscape should be 0 (Figure 9), with equal probabilities of being in each state. To get the free energy land-

scape where the unfolded state is higher in free energy, the probability to convert to the folded state needs to be higher than the probability to convert to the unfolded state.

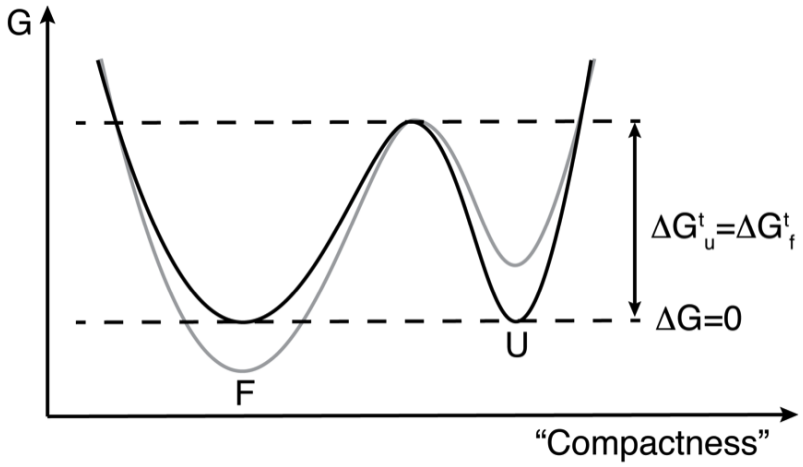


Figure F.9: The free energy landscape shown in Figure 8 is in gray. The free energy landscape where $K_{eq} = 1$ and $\Delta G = 0$ is in black.

How would you get to the following free energy landscape (Figure 10)?

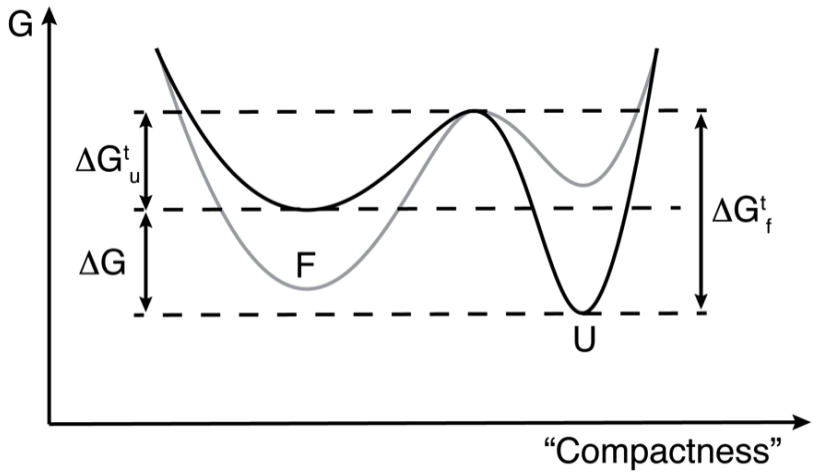


Figure F.10: The free energy landscape shown in Figure 8 is in gray. The desired free energy landscape is in black.

The probability to convert from the folded state to the unfolded state should be greater than the probability to convert from the unfolded state to the folded state.

10. Now set "the probability to convert to the unfolded state if the previous state was folded" equal to 1 and "the probability to convert to the folded state if the previous state was unfolded" equal to zero and run the program. Does the time trace look the way you expected it to look? Why? Have you seen that histogram already? When? Why did the kinetics plot show nothing?

    The protein always stays in the unfolded state. It is like the histogram for the model that is always in the unfolded state. There is nothing in the kinetics plot because the kinetics are for the transition from one state to another and there was no state change here.

11. Now set "the probability to convert to the unfolded state if the previous state was folded" equal to zero and "the probability to convert to the folded state if the previous state was unfolded" equal to 1 and run the program. Does the time trace look the way you expected it to look? Why? Have you seen that histogram already? When? Why did the kinetics plot show nothing?

    This protein always stays in the folded state. It is like the histogram for the model that is always in the folded state. There is nothing in the kinetics plot because the kinetics are for the transition from one state to another and there was no state change here.

12. Now set "the probability to convert to the unfolded state if the previous state was folded" equal to 1 and "the probability to convert to the folded state if the previous state was unfolded" equal to 0.5 and run the program. Does the time trace look the way you expected it to look? Why? Why did the kinetics plot show nothing? Can you rationalize why K_{eq} came out to be $\approx$0.5?

    The time trace looks correct because whenever the protein becomes folded, it immediately returns to unfolded. $K_{eq} \approx 0.5$ because the protein is twice as likely to be in the unfolded state as in the folded state and K_{eq} is defined here as $\frac{[F]}{[U]}$. The kinetics plot measures the unfolding rate, which is fast for all members of the sampled population.

13. Now set "the probability to convert to the unfolded state if the previous state was folded" equal to 0.5 and "the probability to convert to the

folded state if the previous state was unfolded" equal to 1 and run the program. Does the time trace look the way you expected it to look? Why? Why did the kinetics plot show a histogram this time as compared to Step 12? Can you rationalize why K_{eq} came out to be $\sim$2? Write down the time constant that you got.

Now that probability is twice as great that the protein is in the folded state as in the unfolded state in any given measurement. There is a variation in the time required to switch to the unfolded state, which causes the kinetics plot to have various times.

14. Let's play around with kinetics now. Keep "the probability to convert to the folded state if the previous state was unfolded" at 1 since it does not matter anyway and change "the probability to convert to the unfolded state if the previous state was folded" to 0.05 and run the program. Does the time trace look the way you expected it to look? Why? How about the kinetics histogram? Is the time constant larger or smaller than in Step 13? Why?

    There is now a much longer time on average for the protein to switch to the unfolded state, which leads to a larger time constant.

15. Now change "the probability to convert to the unfolded state if the previous state was folded" to 0.9 and run the program. Does the time trace look the way you expected it to look? Why? How about the kinetics histogram? Is the time constant larger or smaller than in Step 14? Step 13? Why?

    Now that the probability to unfold is relatively high, the rate at which it happens will be fast on average, leading to a faster (shorter) time constant.

F.5 Protein folding in the cell

There are many reasons why protein folding inside a cell is different from that in the test tube. One of them – the one we will be focusing on – is macromolecular crowding. There are many molecules inside the cell apart from the protein you might be interested in studying. The density inside the

cell can be 400× that of the density of the buffer in a test tube, which means that there are many macromolecules confined to a small volume. This means that the molecules are sterically hindered to sample the conformations that occupy large extended areas of space, perhaps the ones that have large end-to-end distances. We are going to test a simple model for this crowding and see if it changes your results. The paper clip model for the protein will stay the same. However, instead of dropping the paper clips on an empty flat surface, you will now drop the protein model into a rectangular boundary made out of Styrofoam. The paper frame will serve as a spatial constraint preventing the protein from extending and thus decreasing its "conformational entropy".

F.5.1 Activity 1: An unfolded protein inside a cell

1. Put the Styrofoam frame on a flat surface.
2. Take an unfolded protein model, throw it upwards in the air and watch it fall inside the Styrofoam boundary.
3. If the protein did not fall completely inside the Styrofoam frame, repeat step 2.
4. Remove the Styrofoam frame and measure the end-to-end distance of the shape in which the model ended up. Record the end-to-end distance.
5. Repeat steps 1–4 20 times.

F.5.2 Activity 2: A folded protein inside a cell

1. Repeat the same steps as in Section 5.1 but now use a folded protein model instead.
2. Make sure you record your data.

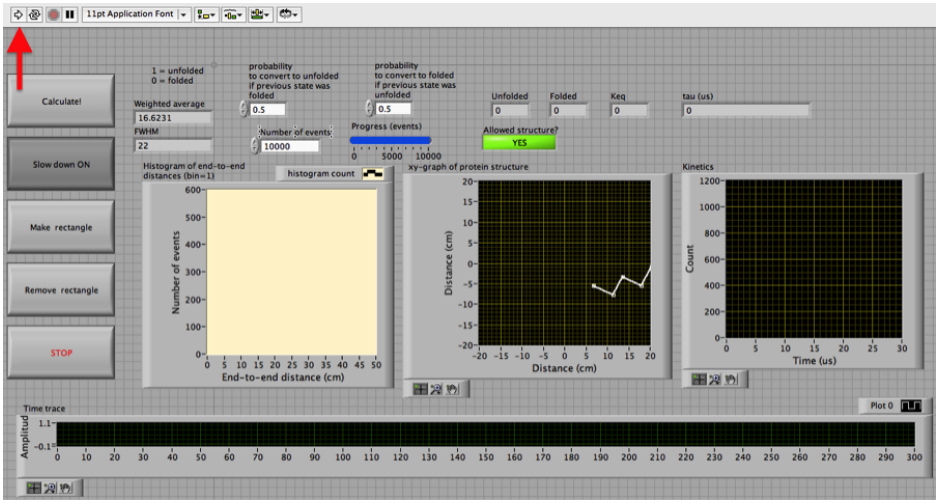
F.5.3 Data analysis

1. Use the data you collected in Sections 5.1 and 5.2 to make two histograms of end-to-end distances for each model, just like you did in Section 2.4.
2. For each histogram calculate the average.
3. Now that you can compare the averages, what can you say about how the histograms are different among the set that you just plotted and also compared to the data you collected previously without the Styrofoam boundary? How would you explain these differences based on the structures that your paper clip models were assuming and the "crowding agent" involved?

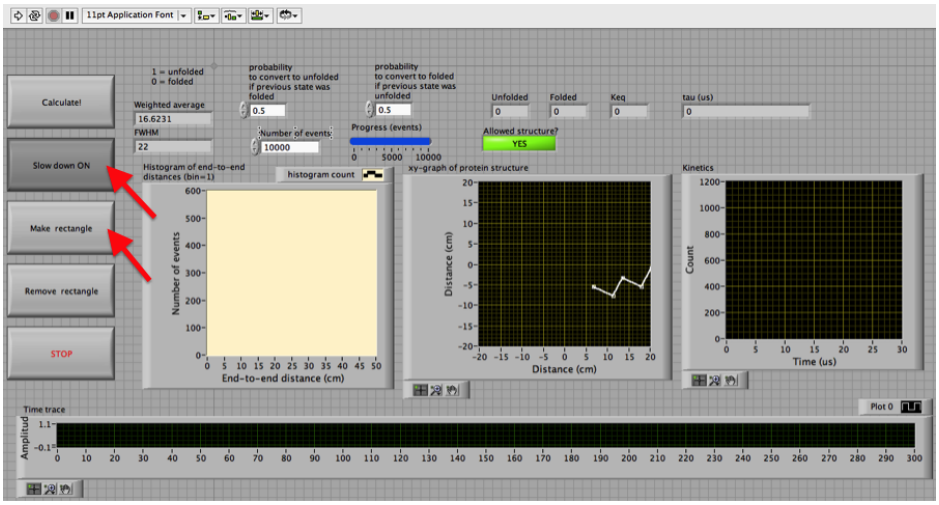
The boundary is expected to decrease the end-to-end distances for both models but the effect is likely to be more pronounced for the folded protein because its end-to-end distance is on overage larger than that of the unfolded protein. In our model, crowding would stabilize the unfolded state by increasing the free energy of the more expanded folded state.

F.5.4 Activity 2: Computer simulation of a crowded protein

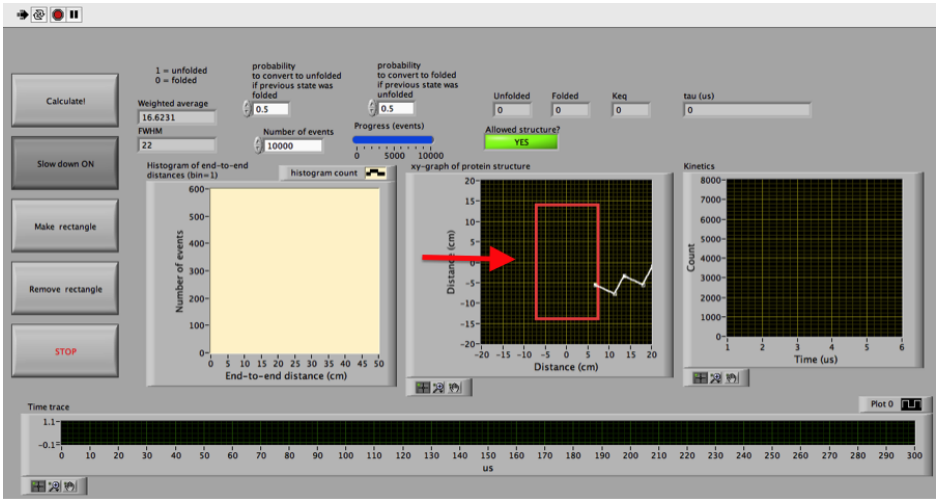
1. The program should already by loaded onto your computer. To start the program press the white arrow on the top left of the screen (highlighted with a red arrow on the figure below).



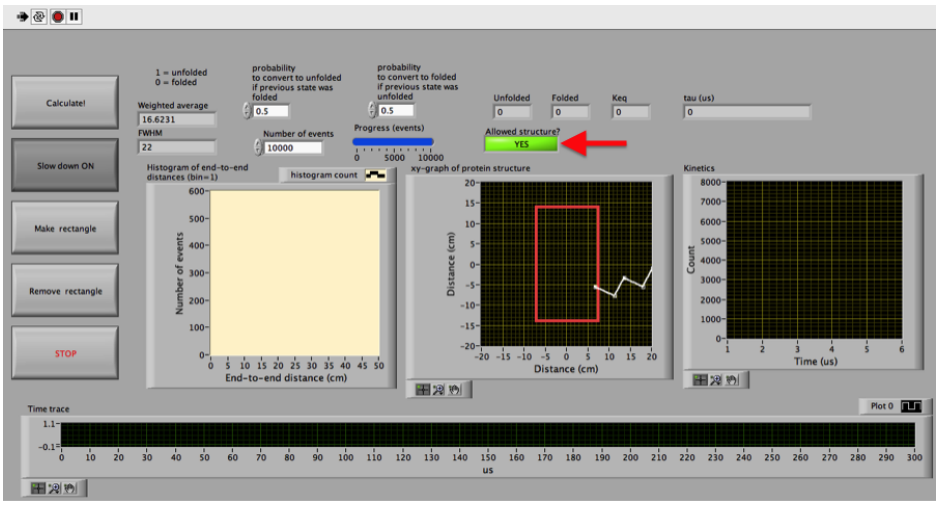
2. Remember those two buttons that I pointed to in Step 7 of Section 4.1 but never explained to you what they did (highlighted with a red arrow on the figure below)?



Press the "Make rectangle" button. You will see a red rectangle in the middle of the graph that displays protein conformations (highlighted with a red arrow on the figure below). If you look closely, that rectangle has dimensions that are identical to the inner rectangle of the Styrofoam frame that you are using. "Remove rectangle" button gets rid of the rectangle if you want to go back to the model without crowding.



3. Make sure you are in the slow mode and press "Calculate!". You will see that even though the protein is cycling through the conformations, the time trace at the bottom is not updated and the "Allowed structure?" indicator (highlighted with a red arrow on the figure below) is dim and reads "NO".



The reason is because the program is set up in such a way that it will

only count the structures that have all 10 of their amino acids inside the red rectangle – the same way as when you did the measurements you only counted the conformations that were completely inside the Styrofoam frame. Turns out that the probability of getting an allowed structure is ∼1 in 60. You can wait for one if you are patient enough and watch "Allowed structure?" indicator light up and read "YES". Otherwise you can just release the "Slow down" button and watch the program do its thing. Of course this time around you have to wait 60× longer for the program to generate 10,000 conformations. You can either wait or stop the program and decrease the number of events to 1,000. You will still get a reasonably accurate result with much less wait time.

When the program is finished running, see what K_{eq} is equal to. Did you expect that based on the results of your measurements? How does this number compare with the results of the simulations without the rectangle (you can quickly repeat that simulation by removing the rectangle with a button on the left panel and running the program again)? Why do you think the two numbers are so different?

$K_{eq} = 0.48$, while without the rectangle $K_{eq} = 1.02$. The unfolded state is favored in the simulation where the boundary is present because in our model the unfolded state can populate more compact structures, which fit inside the boundary more easily.

What effect do you think the crowding constraint would have on protein folding in terms of the shape of the free energy landscape? What state would be more preferred, U or F?

The folded state is likely to go up in free energy, which will make the unfolded state more preferred.

REFERENCES

[1] M. B. Prigozhin and M. Gruebele, "The fast and the slow: Folding and trapping of lambda(6-85)," *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, vol. 133, pp. 19338–19341, Dec 2011.

[2] C. Dumont, Y. Matsumura, S. J. Kim, J. Li, E. Kondrashkina, H. Kihara, and M. Gruebele, "Solvent-tuning the collapse and helix formation time scales of lambda(*)(6-85)," *PROTEIN SCIENCE*, vol. 15, pp. 2596–2604, Nov 2006.

[3] T. Kiefhaber, "Kinetic traps in lysozyme folding," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 92, pp. 9029–9033, Sep 1995.

[4] O. Bieri, G. Wildegger, A. Bachmann, C. Wagner, and T. Kiefhaber, "A salt-induced kinetic intermediate is on a new parallel pathway of lysozyme folding," *BIOCHEMISTRY*, vol. 38, pp. 12460–12470, Sep 1999.

[5] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, "How fast-folding proteins fold," *SCIENCE*, vol. 334, pp. 517–520, Oct 2011.

[6] F. Liu, M. Nakaema, and M. Gruebele, "The transition state transit time of ww domain folding is controlled by energy landscape roughness," *JOURNAL OF CHEMICAL PHYSICS*, vol. 131, p. 195101, Nov 2009.

[7] D. Chandler, *Introduction to modern statistical mechanics*. Oxford University Press, USA, 1987.

[8] W. Y. Yang and M. Gruebele, "Folding at the speed limit," *NATURE*, vol. 423, pp. 193–197, May 2003.

[9] H. R. Ma and M. Gruebele, "Kinetics are probe-dependent during downhill folding of an engineered lambda(6-85) protein," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 102, pp. 2283–2287, Feb 2005.

[10] F. Liu, D. Du, A. A. Fuller, J. E. Davoren, P. Wipf, J. W. Kelly, and M. Gruebele, "An experimental survey of the transition between two-state and downhill protein folding scenarios," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 105, pp. 2369–2374, Feb 2008.

[11] W. Y. Yang and M. Gruebele, "Folding lambda-repressor at its speed limit," *BIOPHYSICAL JOURNAL*, vol. 87, pp. 596–608, Jul 2004.

[12] F. Liu and M. Gruebele, "Tuning lambda(6-85) towards downhill folding at its melting temperature," *JOURNAL OF MOLECULAR BIOLOGY*, vol. 370, pp. 574–584, Jul 2007.

[13] H. R. Ma and M. Gruebele, "Low barrier kinetics: Dependence on observables and free energy surface," *JOURNAL OF COMPUTATIONAL CHEMISTRY*, vol. 27, pp. 125–134, Jan 2006.

[14] P. L. Freddolino, S. Park, B. Roux, and K. Schulten, "Force field bias in protein folding simulations," *BIOPHYSICAL JOURNAL*, vol. 96, pp. 3772–3780, May 2009.

[15] M. S. Shell, R. Ritterson, and K. A. Dill, "A test on peptide stability of amber force fields with implicit solvation," *JOURNAL OF PHYSICAL CHEMISTRY B*, vol. 112, pp. 6878–6886, Jun 2008.

[16] T. Yoda, Y. Sugita, and Y. Okamoto, "Secondary-structure preferences of force fields for proteins evaluated by generalized-ensemble simulations," *CHEMICAL PHYSICS*, vol. 307, pp. 269–283, Dec 2004.

[17] A. Okur, B. Strockbine, V. Hornak, and C. Simmerling, "Using pc clusters to evaluate the transferability of molecular mechanics force fields for proteins," *JOURNAL OF COMPUTATIONAL CHEMISTRY*, vol. 24, pp. 21–31, Jan 2003.

[18] M. Feig, A. D. MacKerell, and C. L. Brooks, "Force field influence on the observation of pi-helical protein structures in molecular dynamics simulations," *JOURNAL OF PHYSICAL CHEMISTRY B*, vol. 107, pp. 2831–2836, Mar 2003.

[19] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, "Evaluation and reparametrization of the opls-aa force field for proteins via comparison with accurate quantum chemical calculations on peptides," *JOURNAL OF PHYSICAL CHEMISTRY B*, vol. 105, pp. 6474–6487, Jul 2001.

[20] W. Damm and W. E. Van Gunsteren, "Reversible peptide folding: Dependence on molecular force field used," *JOURNAL OF COMPUTATIONAL CHEMISTRY*, vol. 21, pp. 774–787, Jul 2000.

[21] K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror, and D. E. Shaw, "Systematic validation of protein force fields against experimental data," *PLOS ONE*, vol. 7, p. e32131, Feb 2012.

[22] S. Piana, K. Lindorff-Larsen, and D. E. Shaw, "How robust are protein folding simulations with respect to force field parameterization? (vol 100, pg l47, 2011)," *BIOPHYSICAL JOURNAL*, vol. 101, pp. 1015–1015, Aug 2011.

[23] Y. Zhu, D. O. V. Alonso, K. Maki, C. Y. Huang, S. J. Lahr, V. Daggett, H. Roder, W. F. DeGrado, and F. Gai, "Ultrafast folding of alpha(3): A de novo designed three-helix bundle protein," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 100, pp. 15486–15491, Dec 2003.

[24] F. Liu, C. Dumont, Y. Zhu, W. F. DeGrado, F. Gai, and M. Gruebele, "A one-dimensional free energy surface does not account for two-probe folding kinetics of protein alpha d-3," *JOURNAL OF CHEMICAL PHYSICS*, vol. 130, p. 061101, Feb 2009.

[25] G. Scott and M. Gruebele, "Solving the low dimensional smoluchowski equation with a singular value basis set," *JOURNAL OF COMPUTATIONAL CHEMISTRY*, vol. 31, pp. 2428–2433, Oct 2010.

[26] K. A. Beauchamp, D. L. Ensign, R. Das, and V. S. Pande, "Quantitative comparison of villin headpiece subdomain simulations and triplet-triplet energy transfer experiments," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 108, pp. 12734–12739, Aug 2011.

[27] T. R. Sosnick, L. Mayne, R. Hiller, and S. W. Englander, "The barriers in protein-folding," *NATURE STRUCTURAL BIOLOGY*, vol. 1, pp. 149–156, Mar 1994.

[28] V. A. Voelz, V. R. Singh, W. J. Wedemeyer, L. J. Lapidus, and V. S. Pande, "Unfolded-state dynamics and structure of protein l characterized by simulation and experiment," *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, vol. 132, pp. 4702–4709, Apr 2010.

[29] F. Liu, Y. G. Gao, and M. Gruebele, "A survey of lambda repressor fragments from two-state to downhill folding," *JOURNAL OF MOLECULAR BIOLOGY*, vol. 397, pp. 789–798, Apr 2010.

[30] W. Y. Yang and M. Gruebele, "Folding at the speed limit," *NATURE*, vol. 423, pp. 193–197, May 2003.

[31] W. Y. Yang and M. Gruebele, "Rate-temperature relationships in lambda-repressor fragment lambda(6-85) folding," *BIOCHEMISTRY*, vol. 43, pp. 13018–13025, Oct 2004.

[32] W. Y. Yang and M. Gruebele, "Folding lambda-repressor at its speed limit," *BIOPHYSICAL JOURNAL*, vol. 87, pp. 596–608, Jul 2004.

[33] G. R. Bowman, V. A. Voelz, and V. S. Pande, "Atomistic folding simulations of the five-helix bundle protein lambda(6-85)," *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, vol. 133, pp. 664–667, Feb 2011.

[34] F. Liu, Y. G. Gao, and M. Gruebele, "A survey of lambda repressor fragments from two-state to downhill folding," *JOURNAL OF MOLECULAR BIOLOGY*, vol. 397, pp. 789–798, Apr 2010.

[35] A. D. Mackerell, "Empirical force fields for biological macromolecules: Overview and issues," *JOURNAL OF COMPUTATIONAL CHEMISTRY*, vol. 25, pp. 1584–1604, Oct 2004.

[36] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, and J. MacKerell, Alexander D., "Optimization of the additive charmm all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles," *JOURNAL OF CHEMICAL THEORY AND COMPUTATION*, vol. 8, pp. 3257–3273, Sep 2012.

[37] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, "Comparison of multiple amber force fields and development of improved protein backbone parameters," *PROTEINS-STRUCTURE FUNCTION AND BIOINFORMATICS*, vol. 65, pp. 712–725, Nov 2006.

[38] F. Liu, Y. G. Gao, and M. Gruebele, "A survey of lambda repressor fragments from two-state to downhill folding," *JOURNAL OF MOLECULAR BIOLOGY*, vol. 397, pp. 789–798, Apr 2010.

[39] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, "How fast-folding proteins fold," *SCIENCE*, vol. 334, pp. 517–520, Oct 2011.

[40] C. B. Anfinsen, E. Haber, M. Sela, and F. H. White, "Kinetics of formation of native ribonuclease during oxidation of reduced polypeptide chain," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 47, no. 9, pp. 1309–1315, 1961.

[41] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips, "3-dimensional model of the myoglobin molecule obtained by x-ray analysis," *NATURE*, vol. 181, no. 4610, pp. 662–666, 1958.

[42] M. Levitt and A. Warshel, "Computer-simulation of protein folding," *NATURE*, vol. 253, no. 5494, pp. 694–698, 1975.

[43] H. Taketomi, Y. Ueda, and N. Go, "Studies on protein folding, unfolding and fluctuations by computer-simulation .1. effect of specific amino-acid sequence represented by specific inter-unit interactions," *INTERNATIONAL JOURNAL OF PEPTIDE AND PROTEIN RESEARCH*, vol. 7, no. 6, pp. 445–459, 1975.

[44] H. Abe and N. Go, "Non-interacting local-structure model of folding and unfolding transition in globular-proteins .2. application to two-dimensional lattice proteins," *BIOPOLYMERS*, vol. 20, no. 5, pp. 1013–1031, 1981.

[45] A. Sali, E. Shakhnovich, and M. Karplus, "Kinetics of protein-folding - a lattice model study of the requirements for folding to the native-state," *JOURNAL OF MOLECULAR BIOLOGY*, vol. 235, pp. 1614–1636, Feb 1994.

[46] Y. Duan and P. A. Kollman, "Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution," *SCIENCE*, vol. 282, pp. 740–744, Oct 1998.

[47] C. M. Phillips, Y. Mizutani, and R. M. Hochstrasser, "Ultrafast thermally-induced unfolding of rnase-a," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 92, pp. 7292–7296, Aug 1995.

[48] G. S. Huang and T. G. Oas, "Submillisecond folding of monomeric lambda-repressor," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 92, pp. 6878–6882, Jul 1995.

[49] A. G. Ladurner, L. S. Itzhaki, V. Daggett, and A. R. Fersht, "Synergy between simulation and experiment in describing the energy landscape of protein folding," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 95, pp. 8473–8478, Jul 1998.

[50] B. Nolting, R. Golbik, and A. R. Fersht, "Submillisecond events in protein-folding," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 92, pp. 10668–10672, Nov 1995.

[51] S. Williams, T. P. Causgrove, R. Gilmanshin, K. S. Fang, R. H. Callender, W. H. Woodruff, and R. B. Dyer, "Fast events in protein folding: Helix melting and formation in a small peptide," *BIOCHEMISTRY*, vol. 35, pp. 691–697, Jan 1996.

[52] R. Gilmanshin, S. Williams, R. H. Callender, W. H. Woodruff, and R. B. Dyer, "Fast events in protein folding: Relaxation dynamics of secondary and tertiary structure in native apomyoglobin," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 94, pp. 3709–3713, Apr 1997.

[53] R. M. Ballew, J. Sabelko, and M. Gruebele, "Observation of distinct nanosecond and microsecond protein folding events," *NATURE STRUCTURAL BIOLOGY*, vol. 3, pp. 923–926, Nov 1996.

[54] R. M. Ballew, J. Sabelko, and M. Gruebele, "Direct observation of fast protein folding: The initial collapse of apomyoglobin," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 93, pp. 5759–5764, Jun 1996.

[55] R. O. Dror, R. M. Dirks, J. . . P. . Grossman, H. Xu, and D. E. Shaw, "Biomolecular simulation: A computational microscope for molecular biology," 2012.

[56] M. Karplus and J. A. McCammon, "Molecular dynamics simulations of biomolecules," *NATURE STRUCTURAL BIOLOGY*, vol. 9, pp. 646–652, Sep 2002.

[57] C. D. Snow, E. J. Sorin, Y. M. Rhee, and V. S. Pande, "How well can simulation predict protein folding kinetics and thermodynamics?," 2005.

[58] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, "Scalable molecular dynamics with namd," *JOURNAL OF COMPUTATIONAL CHEMISTRY*, vol. 26, pp. 1781–1802, Dec 2005.

[59] J. Kubelka, J. Hofrichter, and W. A. Eaton, "The protein folding 'speed limit'," *CURRENT OPINION IN STRUCTURAL BIOLOGY*, vol. 14, pp. 76–88, Feb 2004.

[60] B. Schuler, "Single-molecule fluorescence spectroscopy of protein folding," *CHEMPHYSCHEM*, vol. 6, pp. 1206–1220, Jul 2005.

[61] M. Gruebele, "The fast protein folding problem," *ANNUAL REVIEW OF PHYSICAL CHEMISTRY*, vol. 50, pp. 485–516, 1999.

[62] A. R. Fersht, *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*. W. H. Freeman, 1998.

[63] G. R. Bowman, V. A. Voelz, and V. S. Pande, "Atomistic folding simulations of the five-helix bundle protein lambda(6-85)," *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, vol. 133, pp. 664–667, Feb 2011.

[64] M. B. Prigozhin, K. Sarkar, D. Law, W. C. Swope, M. Gruebele, and J. Pitera, "Reducing lambda repressor to the core," *JOURNAL OF PHYSICAL CHEMISTRY B*, vol. 115, pp. 2090–2096, Mar 2011.

[65] A. Ikai and C. Tanford, "Kinetic evidence for incorrectly folded intermediate states in refolding of denatured proteins," *NATURE*, vol. 230, no. 5289, pp. 100–115, 1971.

[66] J. E. Shea and C. L. Brooks, "From folding theories to folding proteins: A review and assessment of simulation studies of protein folding and unfolding," *ANNUAL REVIEW OF PHYSICAL CHEMISTRY*, vol. 52, pp. 499–535, 2001.

[67] C. D. Snow, N. Nguyen, V. S. Pande, and M. Gruebele, "Absolute comparison of simulated and experimental protein-folding dynamics," *NATURE*, vol. 420, pp. 102–106, Nov 2002.

[68] M. Shirts and V. S. Pande, "Computing - screen savers of the world unite!," *SCIENCE*, vol. 290, pp. 1903–1904, Dec 2000.

[69] Y. Sugita and Y. Okamoto, "Replica-exchange molecular dynamics method for protein folding," *CHEMICAL PHYSICS LETTERS*, vol. 314, pp. 141–151, Nov 1999.

[70] J.-H. Prinz, J. D. Chodera, V. S. Pande, W. C. Swope, J. C. Smith, and F. Noe, "Optimal use of data in parallel tempering simulations for the construction of discrete-state markov models of biomolecular dynamics," *JOURNAL OF CHEMICAL PHYSICS*, vol. 134, p. 244108, Jun 2011.

[71] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, "The energy landscapes and motions of proteins," *SCIENCE*, vol. 254, pp. 1598–1603, Dec 1991.

[72] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, "Funnels, pathways, and the energy landscape of protein-folding - a synthesis," *PROTEINS-STRUCTURE FUNCTION AND GENETICS*, vol. 21, pp. 167–195, Mar 1995.

[73] J. D. Bryngelson and P. G. Wolynes, "Spin-glasses and the statistical-mechanics of protein folding," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 84, pp. 7524–7528, Nov 1987.

[74] J. N. Onuchic, Z. LutheySchulten, and P. G. Wolynes, "Theory of protein folding: The energy landscape perspective," *ANNUAL REVIEW OF PHYSICAL CHEMISTRY*, vol. 48, pp. 545–600, 1997.

[75] K. A. Dill, "Dominant forces in protein folding," *BIOCHEMISTRY*, vol. 29, pp. 7133–7155, Aug 1990.

[76] K. A. Dill and H. S. Chan, "From levinthal to pathways to funnels," *NATURE STRUCTURAL BIOLOGY*, vol. 4, pp. 10–19, Jan 1997.

[77] A. P. Capaldi, M. C. R. Shastry, C. Kleanthous, H. Roder, and S. E. Radford, "Ultrarapid mixing experiments reveal that im7 folds via an on-pathway intermediate," *NATURE STRUCTURAL BIOLOGY*, vol. 8, pp. 68–72, Jan 2001.

[78] S. W. Englander, T. R. Sosnick, L. C. Mayne, M. Shtilerman, P. X. Qi, and Y. W. Bai, "Fast and slow folding in cytochrome c," *ACCOUNTS OF CHEMICAL RESEARCH*, vol. 31, pp. 737–744, Nov 1998.

[79] S. J. Kim, Y. Matsumura, C. Dumont, H. Kihara, and M. Gruebele, "Slowing down downhill folding: A three-probe study," *BIOPHYSICAL JOURNAL*, vol. 97, pp. 295–302, Jul 2009.

[80] M. Gruebele, "Downhill protein folding: evolution meets physics," *COMPTES RENDUS BIOLOGIES*, vol. 328, pp. 701–712, Aug 2005.

[81] D. Tobi and I. Bahar, "Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 102, pp. 18908–18913, Dec 2005.

[82] D. D. Boehr, R. Nussinov, and P. E. Wright, "The role of dynamic conformational ensembles in biomolecular recognition," *NATURE CHEMICAL BIOLOGY*, vol. 5, pp. 789–796, Nov 2009.

[83] M. Gruebele and A. H. Zewail, "Ultrafast reaction dynamics," *PHYSICS TODAY*, vol. 43, pp. 24–33, May 1990.

[84] M. Buscaglia, J. Kubelka, W. A. Eaton, and J. Hofrichter, "Determination of ultrafast protein folding rates from loop formation dynamics," *JOURNAL OF MOLECULAR BIOLOGY*, vol. 347, pp. 657–664, Apr 2005.

[85] M. Buscaglia, L. J. Lapidus, W. A. Eaton, and J. Hofrichter, "Effects of denaturants on the dynamics of loop formation in polypeptides," *BIOPHYSICAL JOURNAL*, vol. 91, pp. 276–288, Jul 2006.

[86] L. J. Lapidus, P. J. Steinbach, W. A. Eaton, A. Szabo, and J. Hofrichter, "Effects of chain stiffness on the dynamics of loop formation in polypeptides. appendix: Testing a 1-dimensional diffusion model for peptide dynamics," *JOURNAL OF PHYSICAL CHEMISTRY B*, vol. 106, pp. 11628–11640, Nov 2002.

[87] R. A. Bertsch, N. Vaidehi, S. I. Chan, and W. A. Goddard, "Kinetic steps for alpha-helix formation," *PROTEINS-STRUCTURE FUNCTION AND BIOINFORMATICS*, vol. 33, pp. 343–357, Nov 1998.

[88] B. Fierz, A. Reiner, and T. Kiefhaber, "Local conformational dynamics in alpha-helices measured by fast triplet transfer," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 106, pp. 1057–1062, Jan 2009.

[89] F. Blanco, M. Ramirez-Alvarado, and L. Serrano, "Formation and stability of beta-hairpin structures in polypeptides," *CURRENT OPINION IN STRUCTURAL BIOLOGY*, vol. 8, pp. 107–111, Feb 1998.

[90] A. R. Dinner, T. Lazaridis, and M. Karplus, "Understanding beta-hairpin formation," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 96, pp. 9068–9073, Aug 1999.

[91] V. Munoz, P. A. Thompson, J. Hofrichter, and W. A. Eaton, "Folding dynamics and mechanism of beta-hairpin formation," *NATURE*, vol. 390, pp. 196–199, Nov 1997.

[92] B. Wathen and Z. Jia, "Protein beta-sheet nucleation is driven by local modular formation," *JOURNAL OF BIOLOGICAL CHEMISTRY*, vol. 285, pp. 18376–18384, Jun 2010.

[93] T. Cellmer, E. R. Henry, J. Hofrichter, and W. A. Eaton, "Measuring internal friction of an ultrafast-folding protein," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 105, pp. 18320–18325, Nov 2008.

[94] R. Narayanan, L. Pelakh, and S. J. Hagen, "Solvent friction changes the folding pathway of the tryptophan zipper tz2," *JOURNAL OF MOLECULAR BIOLOGY*, vol. 390, pp. 538–546, Jul 2009.

[95] J. C. F. Schulz, L. Schmidt, R. B. Best, J. Dzubiella, and R. R. Netz, "Peptide chain dynamics in light and heavy water: Zooming in on

internal friction," *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, vol. 134, pp. 6273–6279, Apr 2012.

[96] L. J. Lapidus, W. A. Eaton, and J. Hofrichter, "Measuring the rate of intramolecular contact formation in polypeptides," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 97, pp. 7220–7225, Jun 2000.

[97] I. C. Yeh and G. Hummer, "Peptide loop-closure kinetics from microsecond molecular dynamics simulations in explicit solvent," *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, vol. 124, pp. 6563–6568, Jun 2002.

[98] H. R. Ma, C. Z. Wan, and A. H. Zewail, "Ultrafast t-jump in water: Studies of conformation and reaction dynamics at the thermal limit," *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, vol. 128, pp. 6338–6340, May 2006.

[99] O. F. Mohammed, G. S. Jas, M. M. Lin, and A. H. Zewail, "Primary peptide folding dynamics observed with ultrafast temperature jump," *ANGEWANDTE CHEMIE-INTERNATIONAL EDITION*, vol. 48, no. 31, pp. 5628–5632, 2009.

[100] D. De Sancho and R. B. Best, "What is the time scale for alpha-helix nucleation?," *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, vol. 133, pp. 6809–6816, May 2011.

[101] P. A. Thompson, W. A. Eaton, and J. Hofrichter, "Laser temperature jump study of the helix reversible arrow coil kinetics of an alanine peptide interpreted with a 'kinetic zipper' model," *BIOCHEMISTRY*, vol. 36, pp. 9200–9210, Jul 1997.

[102] P. A. Thompson, V. Munoz, G. S. Jas, E. R. Henry, W. A. Eaton, and J. Hofrichter, "The helix-coil kinetics of a heteropeptide," *JOURNAL OF PHYSICAL CHEMISTRY B*, vol. 104, pp. 378–389, Jan 2000.

[103] C. M. Davis, S. F. Xiao, D. P. Raeigh, and R. B. Dyer, "Raising the speed limit for beta-hairpin formation," *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, vol. 134, pp. 14476–14482, Sep 2012.

[104] W. Y. Yang and M. Gruebele, "Detection-dependent kinetics as a probe of folding landscape microstructure," *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, vol. 126, pp. 7758–7759, Jun 2004.

[105] C. D. Snow, L. L. Qiu, D. G. Du, F. Gai, S. J. Hagen, and V. S. Pande, "Trp zipper folding kinetics by molecular dynamics

and temperature-jump spectroscopy," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 101, pp. 4077–4082, Mar 2004.

[106] W. Y. Yang, J. W. Pitera, W. C. Swope, and M. Gruebele, "Heterogeneous folding of the trpzip hairpin: Full atom simulation and experiment," *JOURNAL OF MOLECULAR BIOLOGY*, vol. 336, pp. 241–251, Feb 2004.

[107] C. W. Manke and M. C. Williams, "Internal viscosity of polymers and the role of solvent resistance," *MACROMOLECULES*, vol. 18, no. 10, pp. 2045–2051, 1985.

[108] A. Soranno, B. Buchli, D. Nettels, R. R. Cheng, S. Mueller-Spaeth, S. H. Pfeil, A. Hoffmann, E. A. Lipman, D. E. Makarov, and B. Schuler, "Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 109, pp. 17800–17806, Oct 2012.

[109] D. Nettels, I. V. Gopich, A. Hoffmann, and B. Schuler, "Ultrafast dynamics of protein collapse from single-molecule photon statistics," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 104, pp. 2655–2660, Feb 2007.

[110] L. L. Qiu and S. J. Hagen, "A limiting speed for protein folding at low solvent viscosity," *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, vol. 126, pp. 3398–3399, Mar 2004.

[111] L. L. Qiu and S. J. Hagen, "Internal friction in the ultrafast folding of the tryptophan cage," *CHEMICAL PHYSICS*, vol. 307, pp. 243–249, Dec 2004.

[112] B. Zagrovic and V. Pande, "Solvent viscosity dependence of the folding rate of a small protein: Distributed computing study," *JOURNAL OF COMPUTATIONAL CHEMISTRY*, vol. 24, pp. 1432–1436, Sep 2003.

[113] A. Berezhkovskii and A. Szabo, "Time scale separation leads to position-dependent diffusion along a slow coordinate," *JOURNAL OF CHEMICAL PHYSICS*, vol. 135, p. 074108, Aug 2011.

[114] S. J. Hagen, J. Hofrichter, A. Szabo, and W. A. Eaton, "Diffusion-limited contact formation in unfolded cytochrome c: Estimating the maximum rate of protein folding," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 93, pp. 11615–11617, Oct 1996.

[115] J. Sabelko, J. Ervin, and M. Gruebele, "Observation of strange kinetics in protein folding," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 96, pp. 6031–6036, May 1999.

[116] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers, "Atomic-level characterization of the structural dynamics of proteins," *SCIENCE*, vol. 330, pp. 341–346, Oct 2010.

[117] H. S. Chung, J. M. Louis, and W. A. Eaton, "Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 106, pp. 11837–11844, Jul 2009.

[118] T.-H. Lee, L. J. Lapidus, W. Zhao, K. J. Travers, D. Herschlag, and S. Chu, "Measuring the folding transition time of single rna molecules," *BIOPHYSICAL JOURNAL*, vol. 92, pp. 3275–3283, May 2007.

[119] K. Neupane, D. B. Ritchie, H. Yu, D. A. N. Foster, F. Wang, and M. T. Woodside, "Transition path times for nucleic acid folding determined from energy-landscape analysis of single-molecule trajectories," *PHYSICAL REVIEW LETTERS*, vol. 109, p. 068102, Aug 2012.

[120] H. S. Chung, K. McHale, J. M. Louis, and W. A. Eaton, "Single-molecule fluorescence experiments determine protein folding transition path times," *SCIENCE*, vol. 335, pp. 981–984, Feb 2012.

[121] N. T. Southall, K. A. Dill, and A. D. J. Haymet, "A view of the hydrophobic effect," *JOURNAL OF PHYSICAL CHEMISTRY B*, vol. 106, pp. 521–533, Jan 2002.

[122] H. B. Callen, *Thermodynamics and an Introduction to thermostatistics*. Wiley, 1985.

[123] S. Piana, K. Sarkar, K. Lindorff-Larsen, M. Guo, M. Gruebele, and D. E. Shaw, "Computational design and experimental testing of the fastest-folding beta-sheet protein," *JOURNAL OF MOLECULAR BIOLOGY*, vol. 405, pp. 43–48, Jan 2011.

[124] R. B. Best and G. Hummer, "Reaction coordinates and rates from transition paths," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 102, pp. 6732–6737, May 2005.

[125] S. J. DeCamp, A. N. Naganathan, S. A. Waldauer, O. Bakajin, and L. J. Lapidus, "Direct observation of downhill folding of lambda-repressor in a microfluidic mixer," *BIOPHYSICAL JOURNAL*, vol. 97, pp. 1772–1777, Sep 2009.

[126] M. M. Garcia-Mira, M. Sadqi, N. Fischer, J. M. Sanchez-Ruiz, and V. Munoz, "Experimental identification of downhill protein folding," *SCIENCE*, vol. 298, pp. 2191–2195, Dec 2002.

[127] L. L. Qiu, S. A. Pabit, A. E. Roitberg, and S. J. Hagen, "Smaller and faster: The 20-residue trp-cage protein folds in 4 mu s," *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, vol. 124, pp. 12952–12953, Nov 2002.

[128] M. Struthers, J. J. Ottesen, and B. Imperiali, "Design and nmr analyses of compact, independently folded bba motifs," *FOLDING AND DESIGN*, vol. 3, no. 2, pp. 95–103, 1998.

[129] M. R. Bunagan, X. Yang, J. G. Saven, and F. Gai, "Ultrafast folding of a computationally designed trp-cage mutant: Trp(2)-cage," *JOURNAL OF PHYSICAL CHEMISTRY B*, vol. 110, pp. 3759–3763, Mar 2006.

[130] H. Nguyen, M. Jager, J. W. Kelly, and M. Gruebele, "Engineering beta-sheet protein toward the folding speed limit," *JOURNAL OF PHYSICAL CHEMISTRY B*, vol. 109, pp. 15182–15186, Aug 2005.

[131] H. Nguyen, M. Jager, A. Moretto, M. Gruebele, and J. W. Kelly, "Tuning the free-energy landscape of a ww domain by temperature, mutation, and truncation," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 100, pp. 3948–3953, Apr 2003.

[132] M. Jager, H. Nguyen, J. C. Crane, J. W. Kelly, and M. Gruebele, "The folding mechanism of a beta-sheet: The ww domain," *JOURNAL OF MOLECULAR BIOLOGY*, vol. 311, pp. 373–393, Aug 2001.

[133] W. Y. Yang and M. Gruebele, "Rate-temperature relationships in lambda-repressor fragment lambda(6-85) folding," *BIOCHEMISTRY*, vol. 43, pp. 13018–13025, Oct 2004.

[134] S. Deechongkit, H. Nguyen, M. Jager, E. T. Powers, M. Gruebele, and J. W. Kelly, "beta-sheet folding mechanisms from perturbation energetics," *CURRENT OPINION IN STRUCTURAL BIOLOGY*, vol. 16, pp. 94–101, Feb 2006.

[135] M. Jager, Y. Zhang, J. Bieschke, H. Nguyen, M. Dendle, M. E. Bowman, J. P. Noel, M. Gruebele, and J. W. Kelly, "Structure-function-folding relationship in a ww domain," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 103, pp. 10648–10653, Jul 2006.

[136] Y. J. Zhu, X. R. Fu, T. Wang, A. Tamura, S. Takada, J. G. Savan, and F. Gai, "Guiding the search for a protein's maximum rate of folding," *CHEMICAL PHYSICS*, vol. 307, pp. 99–109, Dec 2004.

[137] T. Wang, Y. J. Zhu, and F. Gai, "Folding of a three-helix bundle at the folding speed limit," *JOURNAL OF PHYSICAL CHEMISTRY B*, vol. 108, pp. 3694–3697, Mar 2004.

[138] M. L. Azoitei, Y.-E. A. Ban, J.-P. Julien, S. Bryson, A. Schroeter, O. Kalyuzhniy, J. R. Porter, Y. Adachi, D. Baker, E. F. Pai, and W. R. Schief, "Computational design of high-affinity epitope scaffolds by backbone grafting of a linear epitope," *JOURNAL OF MOLECULAR BIOLOGY*, vol. 415, pp. 175–192, Jan 2012.

[139] J. Rajadas, C. W. Liu, P. Novick, N. W. Kelley, M. Inayathullah, M. C. LeMieux, and V. S. Pande, "Rationally designed turn promoting mutation in the amyloid-beta peptide sequence stabilizes oligomers in solution," *PLOS ONE*, vol. 6, p. e21776, Jul 2011.

[140] P. L. Freddolino and K. Schulten, "Common structural transitions in explicit-solvent simulations of villin headpiece folding," *BIOPHYSICAL JOURNAL*, vol. 97, pp. 2338–2347, Oct 2009.

[141] J. Kubelka, W. A. Eaton, and J. Hofrichter, "Experimental tests of villin subdomain folding simulations," *JOURNAL OF MOLECULAR BIOLOGY*, vol. 329, pp. 625–630, Jun 2003.

[142] B. Zagrovic, C. D. Snow, M. R. Shirts, and V. S. Pande, "Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing (vol 323, pg 927, 2002)," *JOURNAL OF MOLECULAR BIOLOGY*, vol. 324, pp. 1051–1051, Dec 2002.

[143] S. Piana, K. Lindorff-Larsen, and D. E. Shaw, "Protein folding kinetics and thermodynamics from atomistic simulation," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 109, pp. 17845–17850, Oct 2012.

[144] S. Park, A. Y. Lau, and B. Roux, "Computing conformational free energy by deactivated morphing," *JOURNAL OF CHEMICAL PHYSICS*, vol. 129, p. 134102, Oct 2008.

[145] J. Mittal and R. B. Best, "Tackling force-field bias in protein folding simulations: Folding of villin hp35 and pin ww domains in explicit water," *BIOPHYSICAL JOURNAL*, vol. 99, pp. L26–L28, Aug 2010.

[146] B. M. Burmann, S. H. Knauer, A. Sevostyanova, K. Schweimer, R. A. Mooney, R. Landick, I. Artsimovitch, and P. Roesch, "An alpha helix to beta barrel domain switch transforms the transcription factor rfah into a translation factor," *CELL*, vol. 150, pp. 291–303, Jul 2012.

[147] O. M. Becker and M. Karplus, "The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics," *JOURNAL OF CHEMICAL PHYSICS*, vol. 106, pp. 1495–1517, Jan 1997.

[148] P. L. Freddolino, F. Liu, M. Gruebele, and K. Schulten, "Ten-microsecond molecular dynamics simulation of a fast-folding ww domain," *BIOPHYSICAL JOURNAL*, vol. 94, pp. L75–L77, May 2008.

[149] Y. Liu, J. Struempfer, P. L. Freddolino, M. Gruebele, and K. Schulten, "Structural characterization of lambda-repressor folding from all-atom molecular dynamics simulations," *JOURNAL OF PHYSICAL CHEMISTRY LETTERS*, vol. 3, pp. 1117–1123, May 2012.

[150] D. L. Ensign, P. M. Kasson, and V. S. Pande, "Heterogeneity even at the speed limit of folding: Large-scale molecular dynamics study of a fast-folding variant of the villin headpiece," *JOURNAL OF MOLECULAR BIOLOGY*, vol. 374, pp. 806–816, Nov 2007.

[151] V. S. Pande, I. Baker, J. Chapman, S. P. Elmer, S. Khaliq, S. M. Larson, Y. M. Rhee, M. R. Shirts, C. D. Snow, E. J. Sorin, and B. Zagrovic, "Atomistic protein folding simulations on the submillisecond time scale using worldwide distributed computing," *BIOPOLYMERS*, vol. 68, pp. 91–109, Jan 2003.

[152] N. Demirdoven, C. M. Cheatum, H. S. Chung, M. Khalil, J. Knoester, and A. Tokmakoff, "Two-dimensional infrared spectroscopy of antiparallel beta-sheet secondary structure," *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, vol. 126, pp. 7981–7990, Jun 2004.

[153] Z. Ganim and A. Tokmakoff, "Spectral signatures of heterogeneous protein ensembles revealed by md simulations of 2dir spectra," *BIOPHYSICAL JOURNAL*, vol. 91, pp. 2636–2646, Oct 2006.

[154] C. Clementi, P. A. Jennings, and J. N. Onuchic, "How native-state topology affects the folding of dihydrofolate reductase and interleukin-1 beta," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 97, pp. 5871–5876, May 2000.

[155] Y. Levy, P. G. Wolynes, and J. N. Onuchic, "Protein topology determines binding mechanism," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 101, pp. 511–516, Jan 2004.

[156] S. S. Plotkin and J. N. Onuchic, "Investigation of routes and funnels in protein folding by free energy functional methods," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 97, pp. 6509–6514, Jun 2000.

[157] A. G. Murzin, S. E. Brenner, T. Hubbard, and C. Chothia, "Scop - a structural classification of proteins database for the investigation of sequences and structures," *JOURNAL OF MOLECULAR BIOLOGY*, vol. 247, pp. 536–540, Apr 1995.

[158] P. Rogen, "Evaluating protein structure descriptors and tuning gauss integral based descriptors," *JOURNAL OF PHYSICS-CONDENSED MATTER*, vol. 17, pp. S1523–S1538, May 2005.

[159] D. N. Ivankov, S. O. Garbuzynskiy, E. Alm, K. W. Plaxco, D. Baker, and A. V. Finkelstein, "Contact order revisited: Influence of protein size on the folding rate," *PROTEIN SCIENCE*, vol. 12, pp. 2057–2062, Sep 2003.

[160] B. G. Keller, J.-H. Prinz, and F. Noe, "Markov models and dynamical fingerprints: Unraveling the complexity of molecular kinetics," *CHEMICAL PHYSICS*, vol. 396, pp. 92–107, Mar 2012.

[161] F. Noe, "Probability distributions of molecular observables computed from markov models," *JOURNAL OF CHEMICAL PHYSICS*, vol. 128, p. 244103, Jun 2008.

[162] J.-H. Prinz, B. Keller, and F. Noe, "Probing molecular kinetics with markov models: metastable states, transition pathways and spectroscopic observables," *PHYSICAL CHEMISTRY CHEMICAL PHYSICS*, vol. 13, no. 38, pp. 16912–16927, 2011.

[163] G. R. Bowman, D. L. Ensign, and V. S. Pande, "Enhanced modeling via network theory: Adaptive sampling of markov state models," *JOURNAL OF CHEMICAL THEORY AND COMPUTATION*, vol. 6, pp. 787–794, Mar 2010.

[164] G. R. Bowman, X. Huang, and V. S. Pande, "Using generalized ensemble simulations and markov state models to identify conformational states," *METHODS*, vol. 49, pp. 197–201, Oct 2009.

[165] P. Mukherjee, I. Kass, I. Arkin, and M. T. Zanni, "Picosecond dynamics of a membrane protein revealed by 2d ir," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 103, pp. 3528–3533, Mar 2006.

[166] S.-H. Shim, D. B. Strasfeld, Y. L. Ling, and M. T. Zanni, "Automated 2d ir spectroscopy using a mid-ir pulse shaper and application of this technology to the human islet amyloid polypeptide," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 104, pp. 14197–14202, Sep 2007.

[167] Z. Ganim, H. S. Chung, A. W. Smith, L. P. Deflores, K. C. Jones, and A. Tokmakoff, "Amide i two-dimensional infrared spectroscopy of proteins," *ACCOUNTS OF CHEMICAL RESEARCH*, vol. 41, pp. 432–441, Mar 2008.

[168] M. Khalil, N. Demirdoven, and A. Tokmakoff, "Coherent 2d ir spectroscopy: Molecular structure and dynamics in solution," *JOURNAL OF PHYSICAL CHEMISTRY A*, vol. 107, pp. 5258–5279, Jul 2003.

[169] Z. Ahmed, I. A. Beta, A. V. Mikhonin, and S. A. Asher, "Uv-resonance raman thermal unfolding study of trp-cage shows that it is not a simple two-state miniprotein," *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, vol. 127, pp. 10943–10950, Aug 2005.

[170] C. M. Halsey, J. Xiong, O. O. Oshokoya, J. A. Johnson, S. Shinde, J. T. Beatty, G. Ghirlanda, R. D. Jiji, and J. W. Cooley, "Simultaneous observation of peptide backbone lipid solvation and alpha-helical structure by deep-uv resonance raman spectroscopy," *CHEMBIOCHEM*, vol. 12, pp. 2125–2128, Sep 2011.

[171] S. A. Oladepo, K. Xiong, Z. Hong, S. A. Asher, J. Handen, and I. K. Lednev, "Uv resonance raman investigations of peptide and protein structure and dynamics," *CHEMICAL REVIEWS*, vol. 112, pp. 2604–2628, May 2012.

[172] M. Wang and R. D. Jiji, "Resolution of localized small molecule-a beta interactions by deep-ultraviolet resonance raman spectroscopy," *BIOPHYSICAL CHEMISTRY*, vol. 158, pp. 96–103, Oct 2011.

[173] D. Svergun, C. Barberato, and M. H. J. Koch, "Crysol - a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates," *JOURNAL OF APPLIED CRYSTALLOGRAPHY*, vol. 28, pp. 768–773, Dec 1995.

[174] R. B. Best and G. Hummer, "Coordinate-dependent diffusion in protein folding," *PROCEEDINGS OF THE NATIONAL ACADEMY OF*

SCIENCES OF THE UNITED STATES OF AMERICA, vol. 107, pp. 1088–1093, Jan 2010.

[175] P. D. Chowdary and M. Gruebele, "Molecules: What kind of a bag of atoms?," *JOURNAL OF PHYSICAL CHEMISTRY A*, vol. 113, pp. 13139–13143, Nov 2009.

[176] B. A. Shoemaker, J. J. Portman, and P. G. Wolynes, "Speeding molecular recognition by using the folding funnel: The fly-casting mechanism," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 97, pp. 8868–+, Aug 2000.

[177] P. Tompa and A. Fersht, *Structure and function of intrinsically disordered proteins*. Chapman and Hall/CRC, 2010.

[178] T. Bartels, J. G. Choi, and D. J. Selkoe, "alpha-synuclein occurs physiologically as a helically folded tetramer that resists aggregation," *NATURE*, vol. 477, pp. 107–U123, Sep 2011.

[179] W. Y. Yang, E. Larios, and M. Gruebele, "On the extended beta-conformation propensity of polypeptides at high temperature," *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, vol. 125, pp. 16220–16227, Dec 2003.

[180] D. Eliezer, J. Yao, H. J. Dyson, and P. E. Wright, "Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding," *NATURE STRUCTURAL BIOLOGY*, vol. 5, pp. 148–155, Feb 1998.

[181] O. F. Lange, N.-A. Lakomek, C. Fares, G. F. Schroeder, K. F. A. Walter, S. Becker, J. Meiler, H. Grubmueller, C. Griesinger, and B. L. de Groot, "Recognition dynamics up to microseconds revealed from an rdc-derived ubiquitin ensemble in solution," *SCIENCE*, vol. 320, pp. 1471–1475, Jun 2008.

[182] J. Kubelka, T. K. Chiu, D. R. Davies, W. A. Eaton, and J. Hofrichter, "Sub-microsecond protein folding," *JOURNAL OF MOLECULAR BIOLOGY*, vol. 359, pp. 546–553, Jun 2006.

[183] S. A. Waldauer, O. Bakajin, and L. J. Lapidus, "Extremely slow intramolecular diffusion in unfolded protein l," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 107, pp. 13713–13717, Aug 2010.

[184] V. A. Voelz, M. Jaeger, S. Yao, Y. Chen, L. Zhu, S. A. Waldauer, G. R. Bowman, M. Friedrichs, O. Bakajin, L. J. Lapidus, S. Weiss, and V. S.

Pande, "Slow unfolded-state structuring in acyl-coa binding protein folding revealed by simulation and experiment," *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, vol. 134, pp. 12565–12577, Aug 2012.

[185] F. Noe, C. Schuette, E. Vanden-Eijnden, L. Reich, and T. R. Weikl, "Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 106, pp. 19011–19016, Nov 2009.

[186] D. L. Ensign and V. S. Pande, "The fip35 ww domain folds with structural and mechanistic heterogeneity in molecular dynamics simulations," *BIOPHYSICAL JOURNAL*, vol. 96, pp. L53–L55, Apr 2009.

[187] J. T. Vivian and P. R. Callis, "Mechanisms of tryptophan fluorescence shifts in proteins," *BIOPHYSICAL JOURNAL*, vol. 80, pp. 2093–2109, May 2001.

[188] M. H. Hecht, H. C. M. Nelson, and R. T. Sauer, "Mutations in lambda-repressors amino-terminal domain - implications for protein stability and dna-binding," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA-BIOLOGICAL SCIENCES*, vol. 80, no. 9, pp. 2676–2680, 1983.

[189] W. A. Lim, A. Hodel, R. T. Sauer, and F. M. Richards, "The crystal-structure of a mutant protein with altered but improved hydrophobic core packing," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 91, pp. 423–427, Jan 1994.

[190] G. S. Huang and T. G. Oas, "Submillisecond folding of monomeric lambda-repressor," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 92, pp. 6878–6882, Jul 1995.

[191] R. E. Burton, G. S. Huang, M. A. Daugherty, T. L. Calderone, and T. G. Oas, "The energy landscape of a fast-folding protein mapped by ala-¿gly substitutions," *NATURE STRUCTURAL BIOLOGY*, vol. 4, pp. 305–310, Apr 1997.

[192] F. Liu, Y. G. Gao, and M. Gruebele, "A survey of lambda repressor fragments from two-state to downhill folding," *JOURNAL OF MOLECULAR BIOLOGY*, vol. 397, pp. 789–798, Apr 2010.

[193] C. Dumont, Y. Matsumura, S. J. Kim, J. Li, E. Kondrashkina, H. Kihara, and M. Gruebele, "Solvent-tuning the collapse and helix formation time scales of lambda(*)(6-85)," *PROTEIN SCIENCE*, vol. 15, pp. 2596–2604, Nov 2006.

[194] M. H. Hecht, J. M. Sturtevant, and R. T. Sauer, "Effect of single amino-acid replacements on the thermal-stability of the nh2-terminal domain of phage lambda-repressor," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA-BIOLOGICAL SCIENCES*, vol. 81, no. 18, pp. 5685–5689, 1984.

[195] E. Larios, J. W. Pitera, W. C. Swope, and M. Gruebele, "Correlation of early orientational ordering of engineered lambda(6-85) structure with kinetics and thermodynamics," *CHEMICAL PHYSICS*, vol. 323, pp. 45–53, Mar 2006.

[196] J. K. Myers and T. G. Oas, "Contribution of a buried hydrogen bond to lambda repressor folding kinetics," *BIOCHEMISTRY*, vol. 38, pp. 6761–6768, May 1999.

[197] W. Y. Yang, E. Larios, and M. Gruebele, "On the extended beta-conformation propensity of polypeptides at high temperature," *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, vol. 125, pp. 16220–16227, Dec 2003.

[198] W. Y. Yang and M. Gruebele, "Rate-temperature relationships in lambda-repressor fragment lambda(6-85) folding," *BIOCHEMISTRY*, vol. 43, pp. 13018–13025, Oct 2004.

[199] S. Ghaemmaghami, J. M. Word, R. E. Burton, J. S. Richardson, and T. G. Oas, "Folding kinetics of a fluorescent variant of monomeric lambda repressor," *BIOCHEMISTRY*, vol. 37, pp. 9179–9185, Jun 1998.

[200] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, "A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules," *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, vol. 117, pp. 5179–5197, May 1995.

[201] A. Onufriev, D. Bashford, and D. A. Case, "Modification of the generalized born model suitable for macromolecules," *JOURNAL OF PHYSICAL CHEMISTRY B*, vol. 104, pp. 3712–3720, Apr 2000.

[202] M. S. Shell, R. Ritterson, and K. A. Dill, "A test on peptide stability of amber force fields with implicit solvation," *JOURNAL OF PHYSICAL CHEMISTRY B*, vol. 112, pp. 6878–6886, Jun 2008.

[203] H. C. Andersen, "Molecular-dynamics simulations at constant pressure and-or temperature," *JOURNAL OF CHEMICAL PHYSICS*, vol. 72, no. 4, pp. 2384–2393, 1980.

[204] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, "Numerical-integration of cartesian equations of motion of a system with constraints - molecular-dynamics of n-alkanes," *JOURNAL OF COMPUTATIONAL PHYSICS*, vol. 23, no. 3, pp. 327–341, 1977.

[205] J. Ervin, J. Sabelko, and M. Gruebele, "Submicrosecond real-time fluorescence sampling: application to protein folding," *JOURNAL OF PHOTOCHEMISTRY AND PHOTOBIOLOGY B-BIOLOGY*, vol. 54, pp. 1–15, Jan 2000.

[206] L. J. Beamer and C. O. Pabo, "Refined 1.8 angstrom crystal-structure of the lambda-repressor operator complex," *JOURNAL OF MOLECULAR BIOLOGY*, vol. 227, pp. 177–196, Sep 1992.

[207] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The protein data bank," *NUCLEIC ACIDS RESEARCH*, vol. 28, pp. 235–242, Jan 2000.

[208] K. Ghosh and K. A. Dill, "Computing protein stabilities from their chain lengths," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 106, pp. 10649–10654, Jun 2009.

[209] R. E. Burton, J. K. Myers, and T. G. Oas, "Protein folding dynamics: Quantitative comparison between theory and experiment," *BIOCHEMISTRY*, vol. 37, pp. 5337–5343, Apr 1998.

[210] P. A. Jennings and P. E. Wright, "Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin," *SCIENCE*, vol. 262, pp. 892–896, Nov 1993.

[211] Y. J. Xu, L. Mayne, and S. W. Englander, "Evidence for an unfolding and refolding pathway in cytochrome c," *NATURE STRUCTURAL BIOLOGY*, vol. 5, pp. 774–778, Sep 1998.

[212] M. S. Evans, I. M. Sander, and P. L. Clark, "Cotranslational folding promotes beta-helix formation and avoids aggregation in vivo," *JOURNAL OF MOLECULAR BIOLOGY*, vol. 383, pp. 683–692, Nov 2008.

[213] W. Y. Yang and M. Gruebele, "Folding lambda-repressor at its speed limit," *BIOPHYSICAL JOURNAL*, vol. 87, pp. 596–608, Jul 2004.

[214] J. J. Portman, S. Takada, and P. G. Wolynes, "Microscopic theory of protein folding rates. i. fine structure of the free energy profile and folding routes from a variational approach," *JOURNAL OF CHEMICAL PHYSICS*, vol. 114, pp. 5069–5081, Mar 2001.

[215] S. Piana, K. Sarkar, K. Lindorff-Larsen, M. Guo, M. Gruebele, and D. E. Shaw, "Computational design and experimental testing of the fastest-folding beta-sheet protein," *JOURNAL OF MOLECULAR BIOLOGY*, vol. 405, pp. 43–48, Jan 2011.

[216] F. Noe, C. Schuette, E. Vanden-Eijnden, L. Reich, and T. R. Weikl, "Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 106, pp. 19011–19016, Nov 2009.

[217] M. H. Hecht, J. M. Sturtevant, and R. T. Sauer, "Effect of single amino-acid replacements on the thermal-stability of the nh2-terminal domain of phage lambda-repressor," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA-BIOLOGICAL SCIENCES*, vol. 81, no. 18, pp. 5685–5689, 1984.

[218] G. S. Huang and T. G. Oas, "Structure and stability of monomeric lambda-repressor - nmr evidence for 2-state folding," *BIOCHEMISTRY*, vol. 34, pp. 3884–3892, Mar 1995.

[219] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, "How fast-folding proteins fold," *SCIENCE*, vol. 334, pp. 517–520, Oct 2011.

[220] S. J. DeCamp, A. N. Naganathan, S. A. Waldauer, O. Bakajin, and L. J. Lapidus, "Direct observation of downhill folding of lambda-repressor in a microfluidic mixer," *BIOPHYSICAL JOURNAL*, vol. 97, pp. 1772–1777, Sep 2009.

[221] F. Liu, M. Nakaema, and M. Gruebele, "The transition state transit time of ww domain folding is controlled by energy landscape roughness," *JOURNAL OF CHEMICAL PHYSICS*, vol. 131, p. 195101, Nov 2009.

[222] H. R. Ma and M. Gruebele, "Kinetics are probe-dependent during downhill folding of an engineered lambda(6-85) protein," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 102, pp. 2283–2287, Feb 2005.

[223] W. Y. Yang, E. Larios, and M. Gruebele, "On the extended beta-conformation propensity of polypeptides at high temperature,"

JOURNAL OF THE AMERICAN CHEMICAL SOCIETY, vol. 125, pp. 16220–16227, Dec 2003.

[224] W. Y. Yang and M. Gruebele, "Kinetic equivalence of the heat and cold structural transitions of lambda(6-85)," *PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY OF LONDON SERIES A-MATHEMATICAL PHYSICAL AND ENGINEERING SCIENCES*, vol. 363, pp. 565–573, Feb 2005.

[225] M. B. Prigozhin, K. Sarkar, D. Law, W. C. Swope, M. Gruebele, and J. Pitera, "Reducing lambda repressor to the core," *JOURNAL OF PHYSICAL CHEMISTRY B*, vol. 115, pp. 2090–2096, Mar 2011.

[226] P. Chugha and T. G. Oas, "Backbone dynamics of the monomeric lambda repressor denatured state ensemble under nondenaturing conditions," *BIOCHEMISTRY*, vol. 46, pp. 1141–1151, Feb 2007.

[227] M. Oliveberg and P. G. Wolynes, "The experimental survey of protein-folding energy landscapes," *QUARTERLY REVIEWS OF BIOPHYSICS*, vol. 38, pp. 245–288, Aug 2005.

[228] N. C. Fitzkee and G. D. Rose, "Reassessing random-coil statistics in unfolded proteins," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 101, pp. 12497–12502, Aug 2004.

[229] D. Hamada, S. Segawa, and Y. Goto, "Non-native alpha-helical intermediate in the refolding of beta-lactoglobulin, a predominantly beta-sheet protein," *NATURE STRUCTURAL BIOLOGY*, vol. 3, pp. 868–873, Oct 1996.

[230] N. D. Socci, J. N. Onuchic, and P. G. Wolynes, "Protein folding mechanisms and the multidimensional folding funnel," *PROTEINS-STRUCTURE FUNCTION AND GENETICS*, vol. 32, pp. 136–158, Aug 1998.

[231] S. Sarupria, T. Ghosh, A. E. Garcia, and S. Garde, "Studying pressure denaturation of a protein by molecular dynamics simulations," *PROTEINS-STRUCTURE FUNCTION AND BIOINFORMATICS*, vol. 78, pp. 1641–1651, May 2010.

[232] R. Callender and R. B. Dyer, "Probing protein dynamics using temperature jump relaxation spectroscopy," *CURRENT OPINION IN STRUCTURAL BIOLOGY*, vol. 12, pp. 628–633, Oct 2002.

[233] R. M. Clegg, E. L. Elson, and B. W. Maxfield, "New technique for optical observation of kinetics of chemical-reactions perturbed by small pressure changes," *BIOPOLYMERS*, vol. 14, no. 4, pp. 883–887, 1975.

[234] M. Schiewek, M. Krumova, G. Hempel, and A. Blume, "Pressure jump relaxation setup with ir detection and millisecond time resolution," *REVIEW OF SCIENTIFIC INSTRUMENTS*, vol. 78, p. 045101, Apr 2007.

[235] C. Dumont, T. Emilsson, and M. Gruebele, "Reaching the protein folding speed limit with large, sub-microsecond pressure jumps," *NATURE METHODS*, vol. 6, pp. 515–U70, Jul 2009.

[236] J.-B. Rouget, M. A. Schroer, C. Jeworrek, M. Puehse, J.-L. Saldana, Y. Bessin, M. Tolan, D. Barrick, R. Winter, and C. A. Royer, "Unique features of the folding landscape of a repeat protein revealed by pressure perturbation," *BIOPHYSICAL JOURNAL*, vol. 98, pp. 2712–2721, Jun 2010.

[237] M. B. Enright and D. M. Leitner, "Mass fractal dimension and the compactness of proteins," *PHYSICAL REVIEW E*, vol. 71, p. 011912, Jan 2005.

[238] P. D. Chowdary and M. Gruebele, "Molecules: What kind of a bag of atoms?," *JOURNAL OF PHYSICAL CHEMISTRY A*, vol. 113, pp. 13139–13143, Nov 2009.

[239] J. Roche, J. A. Caro, D. R. Norberto, P. Barthe, C. Roumestand, J. L. Schlessman, A. E. Garcia, B. Garcia-Moreno E, and C. A. Royer, "Cavities determine the pressure unfolding of proteins," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 109, pp. 6945–6950, May 2012.

[240] L. Mitra, K. Hata, R. Kono, A. Maeno, D. Isom, J.-B. Rouget, R. Winter, K. Akasaka, B. Garcia-Moreno, and C. A. Royer, "V-i-value analysis: A pressure-based method for mapping the folding transition state ensemble of proteins," *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, vol. 129, pp. 14108–+, Nov 2007.

[241] S. Ghaemmaghami, J. M. Word, R. E. Burton, J. S. Richardson, and T. G. Oas, "Folding kinetics of a fluorescent variant of monomeric lambda repressor," *BIOCHEMISTRY*, vol. 37, pp. 9179–9185, Jun 1998.

[242] W. Y. Yang and M. Gruebele, "Rate-temperature relationships in lambda-repressor fragment lambda(6-85) folding," *BIOCHEMISTRY*, vol. 43, pp. 13018–13025, Oct 2004.

[243] E. Larios, J. W. Pitera, W. C. Swope, and M. Gruebele, "Correlation of early orientational ordering of engineered lambda(6-85) structure with kinetics and thermodynamics," *CHEMICAL PHYSICS*, vol. 323, pp. 45–53, Mar 2006.

[244] D. Paschek, S. Gnanakaran, and A. E. Garcia, "Simulations of the pressure and temperature unfolding of an alpha-helical peptide," *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, vol. 102, pp. 6765–6770, May 2005.

[245] G. Panick, R. Malessa, R. Winter, G. Rapp, K. J. Frye, and C. A. Royer, "Structural characterization of the pressure-denatured state and unfolding/refolding kinetics of staphylococcal nuclease by synchrotron small-angle x-ray scattering and fourier-transform infrared spectroscopy," *JOURNAL OF MOLECULAR BIOLOGY*, vol. 275, pp. 389–402, Jan 1998.

[246] W. Y. Yang, E. Larios, and M. Gruebele, "On the extended beta-conformation propensity of polypeptides at high temperature," *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, vol. 125, pp. 16220–16227, Dec 2003.

[247] C. Dumont, Y. Matsumura, S. J. Kim, J. Li, E. Kondrashkina, H. Kihara, and M. Gruebele, "Solvent-tuning the collapse and helix formation time scales of lambda(*)(6-85)," *PROTEIN SCIENCE*, vol. 15, pp. 2596–2604, Nov 2006.

[248] J. Li, M. Shinjo, Y. Matsumura, M. Morita, D. Baker, M. Ikeguchi, and H. Kihara, "An alpha-helical burst in the src sh3 folding pathway," *BIOCHEMISTRY*, vol. 46, pp. 5072–5082, May 2007.

[249] D. Hamada, S. Segawa, and Y. Goto, "Non-native alpha-helical intermediate in the refolding of beta-lactoglobulin, a predominantly beta-sheet protein," *NATURE STRUCTURAL BIOLOGY*, vol. 3, pp. 868–873, Oct 1996.

[250] P. L. Freddolino, S. Park, B. Roux, and K. Schulten, "Force field bias in protein folding simulations," *BIOPHYSICAL JOURNAL*, vol. 96, pp. 3772–3780, May 2009.

[251] M. B. Prigozhin, K. Sarkar, D. Law, W. C. Swope, M. Gruebele, and J. Pitera, "Reducing lambda repressor to the core," *JOURNAL OF PHYSICAL CHEMISTRY B*, vol. 115, pp. 2090–2096, Mar 2011.

[252] J. Jorgensen, W. L. aand Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *JOURNAL OF CHEMICAL PHYSICS*, vol. 79, no. 2, pp. 926–935, 1983.

[253] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, "Scalable

molecular dynamics with namd," *JOURNAL OF COMPUTATIONAL CHEMISTRY*, vol. 26, pp. 1781–1802, Dec 2005.

[254] D. E. Shaw, M. M. Deneroff, R. O. Dror, J. S. Kuskin, R. H. Larson, J. K. Salmon, C. Young, B. Batson, K. J. Bowers, J. C. Chao, M. P. Eastwood, J. Gagliardo, J. . . P. . Grossman, C. R. Ho, D. J. Ierardi, I. Kolossvary, J. L. Klepeis, T. Layman, C. Mcleavey, M. A. Moraes, R. Mueller, E. C. Priest, Y. Shan, J. Spengler, M. Theobald, B. Towles, and S. C. Wang, "Anton, a special-purpose machine for molecular dynamics simulation," *COMMUNICATIONS OF THE ACM*, vol. 51, pp. 91–97, Jul 2008.

[255] L. J. Beamer and C. O. Pabo, "Refined 1.8 angstrom crystal-structure of the lambda-repressor operator complex," *JOURNAL OF MOLECULAR BIOLOGY*, vol. 227, pp. 177–196, Sep 1992.

[256] G. R. Bowman, V. A. Voelz, and V. S. Pande, "Atomistic folding simulations of the five-helix bundle protein lambda(6-85)," *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, vol. 133, pp. 664–667, Feb 2011.

[257] M. B. Prigozhin and M. Gruebele, "The fast and the slow: Folding and trapping of lambda(6-85)," *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, vol. 133, pp. 19338–19341, Dec 2011.

[258] G. Panick, G. J. A. Vidugiris, R. Malessa, G. Rapp, R. Winter, and C. A. Royer, "Exploring the temperature-pressure phase diagram of staphylococcal nuclease," *BIOCHEMISTRY*, vol. 38, pp. 4157–4164, Mar 1999.

[259] H. Imamura, Y. Isogai, and M. Kato, "Differences in the structural stability and cooperativity between monomeric variants of natural and de novo cro proteins revealed by high-pressure fourier transform infrared spectroscopy," *BIOCHEMISTRY*, vol. 51, pp. 3539–3546, May 2012.

[260] J.-B. Rouget, T. Aksel, J. Roche, J.-L. Saldana, A. E. Garcia, D. Barrick, and C. A. Royer, "Size and sequence and the volume change of protein folding," *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, vol. 133, pp. 6020–6027, Apr 2011.

[261] A. D. Mackerell, "Empirical force fields for biological macromolecules: Overview and issues," *JOURNAL OF COMPUTATIONAL CHEMISTRY*, vol. 25, pp. 1584–1604, Oct 2004.

[262] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux,

M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, "All-atom empirical potential for molecular modeling and dynamics studies of proteins," *JOURNAL OF PHYSICAL CHEMISTRY B*, vol. 102, pp. 3586–3616, Apr 1998.

[263] A. D. MacKerell, M. Feig, and C. L. Brooks, "Improved treatment of the protein backbone in empirical force fields," *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, vol. 126, pp. 698–699, Jan 2004.

[264] S. Piana, K. Lindorff-Larsen, and D. E. Shaw, "How robust are protein folding simulations with respect to force field parameterization?," *BIOPHYSICAL JOURNAL*, vol. 100, pp. L47–L49, May 2011.

[265] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, and J. MacKerell, Alexander D., "Optimization of the additive charmm all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles," *JOURNAL OF CHEMICAL THEORY AND COMPUTATION*, vol. 8, pp. 3257–3273, Sep 2012.

[266] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, "Comparison of multiple amber force fields and development of improved protein backbone parameters," *PROTEINS-STRUCTURE FUNCTION AND BIOINFORMATICS*, vol. 65, pp. 712–725, Nov 2006.

[267] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, "How fast-folding proteins fold," *SCIENCE*, vol. 334, pp. 517–520, Oct 2011.

[268] Y. Liu, J. Struempfer, P. L. Freddolino, M. Gruebele, and K. Schulten, "Structural characterization of lambda-repressor folding from all-atom molecular dynamics simulations," *JOURNAL OF PHYSICAL CHEMISTRY LETTERS*, vol. 3, pp. 1117–1123, May 2012.

[269] S. Piana, K. Sarkar, K. Lindorff-Larsen, M. Guo, M. Gruebele, and D. E. Shaw, "Computational design and experimental testing of the fastest-folding beta-sheet protein," *JOURNAL OF MOLECULAR BIOLOGY*, vol. 405, pp. 43–48, Jan 2011.

[270] C. N. Pace and J. M. Scholtz, "A helix propensity scale based on experimental studies of peptides and proteins," *BIOPHYSICAL JOURNAL*, vol. 75, pp. 422–427, Jul 1998.

[271] R. Coll, B. France, and I. Taylor, "The role of models/and analogies in science education: implications from research," *INTERNATIONAL JOURNAL OF SCIENCE EDUCATION*, vol. 27, pp. 183–198, FEB 4 2005.

[272] A. Harrison and D. Treagust, "A typology of school science models," *INTERNATIONAL JOURNAL OF SCIENCE EDUCATION*, vol. 22, pp. 1011–1026, SEP 2000.

[273] J. Gobert and B. Buckley, "Introduction to model-based teaching and learning in science education," *INTERNATIONAL JOURNAL OF SCIENCE EDUCATION*, vol. 22, pp. 891–894, SEP 2000.

[274] "National research council next generation science standards: For states, by states," 2013.

[275] R. Justi and J. Gilbert, "Modelling, teachers' views on the nature of modelling, and implications for the education of modellers," *INTERNATIONAL JOURNAL OF SCIENCE EDUCATION*, vol. 24, pp. 369–387, APR 2002.

[276] L. GROSSLIGHT, C. UNGER, E. JAY, and C. SMITH, "UNDERSTANDING MODELS AND THEIR USE IN SCIENCE - CONCEPTIONS OF MIDDLE AND HIGH-SCHOOL-STUDENTS AND EXPERTS," *JOURNAL OF RESEARCH IN SCIENCE TEACHING*, vol. 28, pp. 799–822, NOV 1991.

[277] A. G. Harrison and D. F. Treagust, "Modelling in science lessons: Are there better ways to learn with models?," *School Science and Mathematics*, vol. 98, no. 8, pp. 420–429, 1998.

[278] S. Anthony-Cahill, "Using the protein folding literature to teach biophysical chemistry to undergraduates," *BIOCHEMISTRY AND MOLECULAR BIOLOGY EDUCATION*, vol. 29, pp. 45–49, MAR 2001.

[279] R. Bowen, R. Hartung, and Y. Gindt, "A simple protein purification and folding experiment for general chemistry laboratory," *JOURNAL OF CHEMICAL EDUCATION*, vol. 77, pp. 1456–1457, NOV 2000.

[280] C. Jones, "An introduction to research in protein folding for undergraduates," *JOURNAL OF CHEMICAL EDUCATION*, vol. 74, pp. 1306–1310, NOV 1997.

[281] C. Smith, "How do proteins fold?," *BIOCHEMICAL EDUCATION*, vol. 28, pp. 76–79, MAR 2000.

[282] T. Ulmer, A. Bax, N. Cole, and R. Nussbaum, "Structure and dynamics of micelle-bound human alpha-synuclein," *JOURNAL OF BIOLOGICAL CHEMISTRY*, vol. 280, pp. 9595–9603, MAR 11 2005.

[283] T. FORSTER, "*ZWISCHENMOLEKULARE ENERGIEWANDERUNG UND FLUORESZENZ," *ANNALEN DER PHYSIK*, vol. 2, no. 1-2, pp. 55–75, 1948.

[284] S. ZIMMERMAN and S. TRACH, "ESTIMATION OF MACROMOLECULE CONCENTRATIONS AND EXCLUDED VOLUME EFFECTS FOR THE CYTOPLASM OF ESCHERICHIA-COLI," *JOURNAL OF MOLECULAR BIOLOGY*, vol. 222, pp. 599–620, DEC 5 1991.

[285] N. Rutten, W. R. van Joolingen, and J. T. van der Veen, "The learning effects of computer simulations in science education," *COMPUTERS & EDUCATION*, vol. 58, pp. 136–153, JAN 2012.

[286] M. Linn, "Technology and science education: starting points, research programs, and trends," *INTERNATIONAL JOURNAL OF SCIENCE EDUCATION*, vol. 25, pp. 727–758, JUN 2003.

[287] B. White, A. Kahriman, L. Luberice, and F. Idleh, "Evaluation of Software for Introducing Protein Structure VISUALIZATION AND SIMULATION," *BIOCHEMISTRY AND MOLECULAR BIOLOGY EDUCATION*, vol. 38, pp. 284–289, SEP-OCT 2010.

[288] J. Gonzalez-Cruz, R. Rodriguez-Sotres, and M. Rodriguez-Penagos, "On the convenience of using a computer simulation to teach enzyme kinetics to undergraduate students with biological chemistry-related curricula," *BIOCHEMISTRY AND MOLECULAR BIOLOGY EDUCATION*, vol. 31, pp. 93–101, MAR-APR 2003.

[289] K. Sarkar, K. Meister, A. Sethi, and M. Gruebele, "Fast folding of an rna tetraloop on a rugged energy landscape detected by a stacking-sensitive probe," *BIOPHYSICAL JOURNAL*, vol. 97, pp. 1418–1427, Sep 2009.

[290] L. J. Beamer and C. O. Pabo, "Refined 1.8 angstrom crystal-structure of the lambda-repressor operator complex," *JOURNAL OF MOLECULAR BIOLOGY*, vol. 227, pp. 177–196, Sep 1992.

[291] W. Humphrey, A. Dalke, and K. Schulten, "Vmd: Visual molecular dynamics," *JOURNAL OF MOLECULAR GRAPHICS AND MODELLING*, vol. 14, pp. 33–38, Feb 1996.

[292] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, "Scalable molecular dynamics with namd," *JOURNAL OF COMPUTATIONAL CHEMISTRY*, vol. 26, pp. 1781–1802, Dec 2005.

[293] T. Darden, D. York, and L. Pedersen, "Particle mesh ewald - an n.log(n) method for ewald sums in large systems," *JOURNAL OF CHEMICAL PHYSICS*, vol. 98, pp. 10089–10092, Jun 1993.

[294] H. C. Andersen, "Rattle - a velocity version of the shake algorithm for molecular-dynamics calculations," *JOURNAL OF COMPUTATIONAL PHYSICS*, vol. 52, no. 1, pp. 24–34, 1983.

[295] S. Miyamoto and P. A. Kollman, "Settle - an analytical version of the shake and rattle algorithm for rigid water models," *JOURNAL OF COMPUTATIONAL CHEMISTRY*, vol. 13, pp. 952–962, Oct 1992.

[296] D. E. Shaw, M. M. Deneroff, R. O. Dror, J. S. Kuskin, R. H. Larson, J. K. Salmon, C. Young, B. Batson, K. J. Bowers, J. C. Chao, M. P. Eastwood, J. Gagliardo, J. P. Grossman, C. R. Ho, D. J. Ierardi, I. Kolossvary, J. L. Klepeis, T. Layman, C. Mcleavey, M. A. Moraes, R. Mueller, E. C. Priest, Y. B. Shan, J. Spengler, M. Theobald, B. Towles, and S. C. Wang, "Anton, a special-purpose machine for molecular dynamics simulation," *COMMUNICATIONS OF THE ACM*, vol. 51, pp. 91–97, Jul 2008.